PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C07K

A2

(11) International Publication Number: **WO 00/69885**

(43) International Publication Date: 23 November 2000 (23.11.00)

(21) International Application Number: PCT/US00/13361

(22) International Filing Date: 15 May 2000 (15.05.00)

(30) Priority Data:
09/312,359 14 May 1999 (14.05.99) US

(71) Applicant: MILLENNIUM PHARMACEUTICALS, INC. [US/US]; 75 Sidney Street, Cambridge, MA 02139 (US).

(72) Inventors: PAN, Yang; 6 Hamilton Road #1, Brookline, MA 02146 (US). LEIBY, Kevin, R.; 4 Skowhegan Way, Natick, MA 01760 (US).

(74) Agents: CORUZZI, Laura, A. et al.; Pennie & Edmonds LLP, 1155 Avenue of the Americas, New York, NY 10036 (US).

(81) Designated States: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: SECRETED PROTEINS AND USES THEREOF

```
GTCGACCCACGCGTCCGCTCCCGGGCCGCGGCCAGCCGAGGATCAGAGCGCTTGCCCCGGGGGTCTGTCGCGTCGCAGG      79

                    M  R  T  L  W  M  A  L  C  A  L  S  R  L  W      15
CGCCGCCGGAGCCCTGGCC ATG AGG ACC CTG TGG ATG GCG CTG TGC GCG CTG TCG CGG CTG TGG      143

 P  G  A  Q  A  G  C  A  E  A  G  R  C  C  P  G  R  D  P  A      35
CCC GGG GCC CAG GCC GGC TGC GCC GAG GCC GGG CGC TGC TGT CCC GGC CGG GAC CCC GCC      203

 C  F  A  R  G  W  R  L  D  R  V  Y  G  T  C  F  C  D  Q  A      55
TGC TTC GCC CGC GGC TGG AGG CTG GAC AGG GTC TAC GGG ACG TGT TTC TGC GAC CAA GCC      263

 C  R  F  T  G  D  C  C  F  D  Y  D  R  A  C  P  A  R  P  C      75
TGT CGC TTC ACC GGG GAC TGC TGC TTC GAC TAC GAC AGG GCG TGC CCA GCT CGC CCG TGC      323

 F  V  G  E  W  S  P  W  S  G  C  A  D  Q  C  K  P  T  T  R      95
TTC GTG GGG GAA TGG AGC CCC TGG AGT GGT TGT GCA GAC CAG TGC AAG CCT ACA ACC CGT      383

 V  R  R  R  S  V  Q  Q  E  P  Q  N  G  G  A  P  C  P  P  L     115
GTG CGG AGG CGC TCG GTG CAG CAG GAG CCT CAG AAC GGC GGG GCG CCC TGC CCA CCC CTG      443

 E  E  R  A  G  C  L  E  Y  S  T  P  Q  G  Q  D  C  G  H  T     135
GAA GAG AGA GCT GGC TGC CTG GAG TAC TCC ACC CCG CAG GGC CAG GAC TGC GGG CAC ACC      503

 Y  V  P  A  F  I  T  T  S  A  F  N  K  E  R  T  R  Q  A  T     155
TAT GTT CCT GCC TTT ATA ACT ACC TCC GCA TTC AAC AAG GAG AGA ACA CGA CAA GCT ACG      563

 S  P  H  W  S  T  H  T  E  D  A  G  Y  C  M  E  F  K  T  E     175
TCT CCA CAC TGG TCT ACA CAC ACA GAG GAT GCT GGA TAC TGT ATG GAG TTT AAG ACA GAG      623
```

(57) Abstract

The invention provides isolated nucleic acid molecules, designated TANGO 204, TANGO 206, TANGO 209, and A236, which encode wholly secreted or membrane-associated proteins. The invention also provides antisense nucleic acid molecules, expression vectors containing the nucleic acid molecules of the invention, host cells into which the expression vectors have been introduced, and non-human transgenic animals in which a nucleic acid molecule of the invention has been introduced or disrupted. The invention still further provides isolated polypeptides, fusion polypeptides, antigenic peptides and antibodies. Diagnostic, screening and therapeutic methods utilizing compositions of the invention are also provided.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

SECRETED PROTEINS AND USES THEREOF

Cross Reference to Related Applications

This is a continuation-in-part application of co-pending United States Patent Application No. 09/312,359, filed May 14, 1999, the entire contents of which are incorporated herein by reference in its entirety.

Background of the Invention

Many secreted proteins, for example, cytokines, play a vital role in the regulation of cell growth, cell differentiation, and a variety of specific cellular responses. A number of medically useful proteins, including erythropoietin, granulocyte-macrophage colony stimulating factor, human growth hormone, and various interleukins, are secreted proteins. Thus, an important goal in the design and development of new therapies is the identification and characterization of membrane-associated and secreted proteins and the genes which encode them.

Many membrane-associated proteins are receptors which bind a ligand and transduce an intracellular signal, leading to a variety of cellular responses. The identification and characterization of such a receptor enables one to identify both the ligands which bind to the receptor and the intracellular molecules and signal transduction pathways associated with the receptor, permitting one to identify or design modulators of receptor activity, e.g., receptor agonists or antagonists and modulators of signal transduction.

Summary of the Invention

The present invention is based, at least in part, on the discovery of cDNA molecules encoding TANGO 204, TANGO 206, TANGO 209, and A236, all of which are predicted to be either wholly secreted or transmembrane proteins. These proteins, fragments, derivatives, and variants thereof are collectively referred to as a "polypeptides of the invention" or "proteins of the invention." Nucleic acid molecules encoding the polypeptides or proteins of the invention are collectively referred to as "nucleic acids of the invention."

The nucleic acids and polypeptides of the present invention are useful as modulating agents in regulating a variety of cellular processes. Accordingly, in one aspect, this invention provides isolated nucleic acid molecules encoding a polypeptide of the invention or a biologically active portion thereof. The present invention also provides nucleic acid molecules which are suitable for use as primers or hybridization probes for the detection of nucleic acids encoding a polypeptide of the invention.

The invention features nucleic acid molecules which are at least 45% (or 55%, 65%, 75%, 85%, 95%, or 98%) identical to the nucleotide sequence of SEQ ID Nos. 1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 70, 72, 74, 84, 86, 88, 90, 96, 98, 100, 109, 111, 113, 122, 124, 126, 132, 134, 136, 142, 144, 146, 156, 159, 161, or 163, the nucleotide sequence of the cDNA insert of a clone deposited with ATCC® as Accession Number 207192 (the "cDNA of ATCC® 207192"), the nucleotide sequence of the cDNA insert of a clone deposited with ATCC® as Accession Number 207223 (the "cDNA of ATCC® 207223"), the nucleotide sequence of the cDNA insert of a clone deposited with ATCC® as Accession Number PTA-34 (the "cDNA of ATCC® PTA-34"), the nucleotide sequence of the cDNA insert of a clone deposited with ATCC® as Accession Number 207221 (the "cDNA of ATCC® 207221"), or the nucleotide sequence of the cDNA insert of a clone deposited with ATCC® as Accession Number 207189 (the "cDNA of ATCC® 207189"), or a complement thereof.

The invention features nucleic acid molecules which are at least 45% (or 55%, 65%, 75%, 85%, 95%, or 98%) identical to the nucleotide sequence of SEQ ID Nos. 1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 70, 72, 74, 84, 86, 88, 90, 96, 98, 100, 109, 111, 113, 122, 124, 126, 132, 134, 136, 142, 144, 146, 156, 159, 161, or 163, the nucleotide sequence of the cDNA insert of a clone deposited with ATCC® as Accession Number 207192 (the "cDNA of ATCC® 207192"), the nucleotide sequence of the cDNA insert of a clone deposited with ATCC® as Accession Number 207223 (the "cDNA of ATCC® 207223"), the nucleotide sequence of the cDNA insert of a clone deposited with ATCC® as Accession Number PTA-34 (the "cDNA of ATCC® PTA-34"), the nucleotide sequence of the cDNA insert of a clone deposited with ATCC® as Accession Number 207221 (the "cDNA of ATCC® 207221"), or the nucleotide sequence of the cDNA insert of a clone deposited with ATCC® as Accession Number 207189 (the "cDNA of ATCC® 207189"), or a complement thereof, wherein such nucleic acid molecules encode polypeptides or

proteins that exhibit at least one structural and/or functional feature of a polypeptide of the invention.

The invention features nucleic acid molecules comprising at least 25, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1550, 1600, 1650, 1700, 1750, 1800, 1850, 1900, 1950, 2000, 2050, 2100, 2150, 2200, 2250, 2300, 2350, 2400, 2450, 2500, 2550, 2600, 2650, 2700, 2750, 2800, 2850, 2900, 2950, 3000, or 3050 nucleotides of the nucleotide sequence of SEQ ID NO:1, the nucleotide sequence of the human TANGO 204 cDNA of ATCC® 207192, or a complement thereof.

The invention features nucleic acid molecules comprising at least 25, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, or 790 nucleotides of the nucleotide sequence of SEQ ID NO:3, or a complement thereof.

The invention features nucleic acid molecules comprising at least 25, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, or 1290 nucleotides of the nucleotide sequence of SEQ ID NO:13, the nucleotide sequence of the mouse TANGO 204 cDNA of ATCC® 207189, or a complement thereof.

The invention features nucleic acid molecules comprising at least 25, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, or 790 nucleotides of the nucleotide sequence of SEQ ID NO:15, or a complement thereof.

The invention features nucleic acid molecules comprising at least 25, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1550, 1600, 1650, 1700, 1750, 1800, or 1840 nucleotides of the nucleotide sequence of SEQ ID NO:4, the nucleotide sequence of the human TANGO 206 cDNA of ATCC® 207223, or a complement thereof.

The invention features nucleic acid molecules comprising at least 25, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, or 1250 nucleotides of the nucleotide sequence of SEQ ID NO:6, or a complement thereof.

The invention features nucleic acid molecules comprising at least 25, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 650, 700, 750, 800, 850, 900, 950, 1000,

1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1550, 1600, 1650, 1700, 1750, 1800, 1850, 1900, 1950, 2000, 2050, or 2090 nucleotides of the nucleotide sequence of SEQ ID NO:16, the nucleotide sequence of the mouse TANGO 206 cDNA of ATCC® 207221, or a complement thereof.

The invention features nucleic acid molecules comprising at least 25, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, or 1250 or nucleotides of the nucleotide sequence of SEQ ID NO:18, or a complement thereof.

The invention features nucleic acid molecules comprising at least 25, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1550, 1600, 1650, 1700, 1750, 1800, 1850, 1900, 1950, 2000, 2050, 2100, 2150, 2200, 2250, 2300, 2350, 2400, 2450, 2500, 2550, 2600, 2650, 2700, 2750, 2800, 2850, 2900, 2950, 3000, or 3010 nucleotides of the nucleotide sequence of SEQ ID NO:7, the nucleotide sequence of the human TANGO 209 cDNA of ATCC® 207223, or a complement thereof.

The invention features nucleic acid molecules comprising at least 25, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300 or 1335 nucleotides of the nucleotide sequence of SEQ ID NO:9, or a complement thereof.

The invention features nucleic acid molecules comprising at least 25, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1550, 1600, 1650, 1700, 1750, 1800, 1850, 1900, 1950, 2000, 2050, 2100, 2150, 2200, 2250, 2300, 2350, 2400, 2450, 2500, 2550, 2600, 2650, 2700, 2750, 2800, or 2810 nucleotides of the nucleotide sequence of SEQ ID NO:19, the nucleotide sequence of the mouse TANGO 209 cDNA of ATCC® 207221, or a complement thereof.

The invention features nucleic acid molecules comprising at least 25, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300 or 1340 nucleotides of the nucleotide sequence of SEQ ID NO:21, or a complement thereof.

The invention features nucleic acid molecules comprising at least 25, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 650, 700, 750, 800, 850, 900, 950, 1000,

1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1550, 1600, 1650, 1700, 1750, 1800, 1850, 1900, or 1948 nucleotides of the nucleotide sequence of SEQ ID NO:10, the nucleotide sequence of the human TANGO A236 cDNA of ATCC® PTA-34, or a complement thereof.

The invention features nucleic acid molecules comprising at least 25, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, or 1118 nucleotides of the nucleotide sequence of SEQ ID NO:12, or a complement thereof.

The invention features nucleic acid molecules comprising at least 25, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, 1150, 1200, 1250, 1300, 1350, 1400, 1450, 1500, 1550, 1600, 1650, 1700, 1750, 1800, 1850, 1900, or 1945 nucleotides of the nucleotide sequence of the mouse A236 cDNA of SEQ ID NO:156, or a complement thereof.

The invention features nucleic acid molecules comprising at least 25, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1050, 1100, or 1115 nucleotides of the nucleotide sequence of SEQ ID NO:158, or a complement thereof.

The invention features nucleic acid molecules which include a fragment of at least 300 (325, 350, 375, 400, 425, 450, 500, 550, 600, 650, 700, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000, 2400, 2600, 2800, 3000, or 3200) nucleotides of the nucleotide sequence of SEQ ID NO:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, or 156, the nucleotide sequence of the cDNA of ATCC® 207192, the nucleotide sequence of the cDNA of ATCC® 207223, the nucleotide sequence of the cDNA of ATCC® PTA-34, the nucleotide sequence of the cDNA of ATCC® 207189, or the nucleotide sequence of the cDNA of ATCC® 207221, or a complement thereof.

The invention features nucleic acid molecules which include a fragment of at least 300 (325, 350, 375, 400, 425, 450, 500, 550, 600, 650, 700, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000, 2400, 2600, 2800, 3000, or 3200) nucleotides of the nucleotide sequence of SEQ ID Nos. 1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 70, 72, 74, 84, 86, 88, 90, 96, 98, 100, 109, 111, 113, 122, 124, 126, 132, 134, 136, 142, 144, 146, 156, 159, 161, or 163, the nucleotide sequence of the cDNA of ATCC® 207192, the nucleotide sequence of the cDNA of ATCC® 207223, the nucleotide sequence of the cDNA of

ATCC® PTA-34, the nucleotide sequence of the cDNA of ATCC® 207189, or the nucleotide sequence of the cDNA of ATCC® 207221, or a complement thereof, wherein such nucleic acid molecules encode polypeptides or proteins that exhibit at least one structural and/or functional feature of a polypeptide of the invention.

The invention also features nucleic acid molecules which include a nucleotide sequence encoding a protein having an amino acid sequence that is at least 45% (or 55%, 65%, 75%, 85%, 95%, or 98%) identical to the amino acid sequence of SEQ ID Nos. 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 12, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164, the amino acid sequence encoded by the cDNA of ATCC® 207192, the amino acid sequence encoded by the cDNA of ATCC® 207223, the amino acid sequence encoded by the cDNA of ATCC® PTA-34, the amino acid sequence encoded by the cDNA of ATCC® 207189, or the amino acid sequence encoded by the cDNA of ATCC® 207221, or a complement thereof.

The invention also features nucleic acid molecules which include a nucleotide sequence encoding a protein having an amino acid sequence that is at least 45% (or 55%, 65%, 75%, 85%, 95%, or 98%) identical to the amino acid sequence of SEQ ID Nos. 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 12, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164, the amino acid sequence encoded by the cDNA of ATCC® 207192, the amino acid sequence encoded by the cDNA of ATCC® 207223, the amino acid sequence encoded by the cDNA of ATCC® PTA-34, the amino acid sequence encoded by the cDNA of ATCC® 207189, or the amino acid sequence encoded by the cDNA of ATCC® 207221, or a complement thereof, wherein the protein encoded by the nucleotide sequence also exhibits at least one structural and/or functional feature of a polypeptide of the invention.

In one embodiment, the nucleic acid molecules have the nucleotide sequence of SEQ ID NO: 1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, or 156, the nucleotide sequence of the cDNA of ATCC® 207192, the nucleotide sequence of the cDNA of ATCC® 207223, the nucleotide sequence of the cDNA of ATCC® PTA-34, the nucleotide sequence of the cDNA of ATCC® 207189, or the nucleotide sequence of the cDNA of ATCC® 207221, or a complement thereof.

Also within the invention are nucleic acid molecules which encode a fragment of a polypeptide having the amino acid sequence of SEQ ID Nos. 2, 5, 8, 11, 14, 17, 20,

71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 12, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164, or a fragment including at least 15 (20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390 or 400) contiguous amino acids of SEQ ID NO:2, 5, 8, 11, 14, 17, 20, or 157, the amino acid sequence encoded by the cDNA of ATCC® 207192, the amino acid sequence encoded by the cDNA of ATCC® 207223, the amino acid sequence encoded by the cDNA of ATCC® PTA-34, the amino acid sequence encoded by the cDNA of ATCC® 207189, or the amino acid sequence encoded by the cDNA of ATCC® 207221.

Also within the invention are nucleic acid molecules which encode a fragment of a polypeptide having the amino acid sequence of SEQ ID Nos. 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 12, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164, or a fragment including at least 15 (20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390 or 400) contiguous amino acids of SEQ ID Nos. 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 12, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164, the amino acid sequence encoded by the cDNA of ATCC® 207192, the amino acid sequence encoded by the cDNA of ATCC® 207223, the amino acid sequence encoded by the cDNA of ATCC® PTA-34, the amino acid sequence encoded by the cDNA of ATCC® 207189, or the amino acid sequence encoded by the cDNA of ATCC® 207221, wherein the fragment exhibits at least one structural and/or functional feature of a polypeptide of the invention.

The invention includes nucleic acid molecules which encode a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID Nos. 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 12, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164, the amino acid sequence encoded by the cDNA of ATCC® 207192, the amino acid sequence encoded by the cDNA of ATCC® 207223, the amino acid sequence encoded by the cDNA of ATCC® PTA-34, the amino acid sequence encoded by the cDNA of ATCC® 207189, or the amino acid sequence encoded by the cDNA of ATCC® 207221, wherein the nucleic acid molecule hybridizes to a nucleic acid molecule consisting of a nucleic acid sequence encoding SEQ

ID Nos. 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 12, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164, the nucleotide sequence of the cDNA of ATCC® 207192, the nucleotide sequence of the cDNA of ATCC® 207223, the nucleotide sequence of the cDNA of ATCC® PTA-34, the nucleotide sequence of the cDNA of ATCC® 207189, or the nucleotide sequence of the cDNA of ATCC® 207221, or a complement thereof under stringent conditions.

The invention includes nucleic acid molecules which encode a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID Nos. 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 12, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164, the amino acid sequence encoded by the cDNA of ATCC® 207192, the amino acid sequence encoded by the cDNA of ATCC® 207223, the amino acid sequence encoded by the cDNA of ATCC® PTA-34, the amino acid sequence encoded by the cDNA of ATCC® 207189, or the amino acid sequence encoded by the cDNA of ATCC® 207221, wherein the nucleic acid molecule hybridizes to a nucleic acid molecule consisting of a nucleic acid sequence encoding SEQ ID Nos. 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 12, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164, the nucleotide sequence of the cDNA of ATCC® 207192, the nucleotide sequence of the cDNA of ATCC® 207223, the nucleotide sequence of the cDNA of ATCC® PTA-34, the nucleotide sequence of the cDNA of ATCC® 207189, or the nucleotide sequence of the cDNA of ATCC® 207221, or a complement thereof under stringent conditions, wherein such nucleic acid molecules encode polypeptides or proteins that exhibit at least one structural and/or functional feature of a polypeptide of the invention.

Also within the invention are isolated polypeptides or proteins having an amino acid sequence that is at least about 60%, preferably 65%, 75%, 85%, 95%, or 98% identical to the amino acid sequence of SEQ ID Nos. 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 12, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164, the amino acid sequence encoded by the cDNA of ATCC® 207192, the amino acid sequence encoded by the cDNA of ATCC® 207223, the amino acid sequence encoded by the cDNA of ATCC® PTA-34, the amino acid sequence encoded by the cDNA of ATCC® 207189, or the amino acid sequence encoded by the cDNA of ATCC® 207221.

Also within the invention are isolated polypeptides or proteins having an amino acid sequence that is at least about 60%, preferably 65%, 75%, 85%, 95%, or 98% identical to the amino acid sequence of SEQ ID Nos. 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 12, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164, the amino acid sequence encoded by the cDNA of ATCC® 207192, the amino acid sequence encoded by the cDNA of ATCC® 207223, the amino acid sequence encoded by the cDNA of ATCC® PTA-34, the amino acid sequence encoded by the cDNA of ATCC® 207189, or the amino acid sequence encoded by the cDNA of ATCC® 207221, wherein the polypeptides or proteins also exhibit at least one structural and/or functional feature of a polypeptide of the invention.

Also within the invention are isolated polypeptides or proteins which are encoded by a nucleic acid molecule having a nucleotide sequence that is at least about 60%, preferably 65%, 75%, 85%, or 95% identical to the nucleic acid sequence encoding SEQ ID Nos. 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 12, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164, and isolated polypeptides or proteins which are encoded by a nucleic acid molecule having a nucleotide sequence which hybridizes under stringent hybridization conditions to a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 70, 72, 74, 84, 86, 88, 90, 96, 98, 100, 109, 111, 113, 122, 124, 126, 132, 134, 136, 142, 144, 146, 156, 159, 161, or 163, or complement thereof, the non-coding strand of the cDNA of ATCC® 207192, the non-coding strand of the cDNA of ATCC® 207223, the non-coding strand of the cDNA of ATCC® PTA-34, the non-coding strand of the cDNA of ATCC® 207189, or the non-coding strand of the cDNA of ATCC® 207221.

Also within the invention are polypeptides which are naturally occurring allelic variants of a polypeptide that includes the amino acid sequence of SEQ ID Nos. 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 12, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164, the amino acid sequence encoded by the cDNA of ATCC® 207192, the amino acid sequence encoded by the cDNA of ATCC® 207223, the amino acid sequence encoded by the cDNA of ATCC® PTA-34, the amino acid sequence encoded by the cDNA of ATCC® 207189, or the amino acid sequence encoded by the cDNA of ATCC® 207221, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule having the sequence of SEQ ID

NO:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 70, 72, 74, 84, 86, 88, 90, 96, 98, 100, 109, 111, 113, 122, 124, 126, 132, 134, 136, 142, 144, 146, 156, 159, 161, or 163, or a complement thereof, under stringent conditions. Such allelic variant differ at 1%, 2%, 3%, 4%, or 5% of the amino acid residues.

The invention also features nucleic acid molecules that hybridize under stringent conditions to a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 70, 72, 74, 84, 86, 88, 90, 96, 98, 100, 109, 111, 113, 122, 124, 126, 132, 134, 136, 142, 144, 146, 156, 159, 161, or 163, the cDNA of ATCC® 207192, the cDNA of ATCC® 207223, the cDNA of ATCC® PTA-34, the cDNA of 207189, or the cDNA of ATCC® 207221, or a complement thereof. In other embodiments, the nucleic acid molecules are at least 300 (325, 350, 375, 400, 425, 450, 500, 550, 600, 650, 700, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000, 2200, 2400, 2600, 2800, 3000, or 3200) nucleotides in length and hybridize under stringent conditions to a nucleic acid molecule consisting of the nucleotide sequence of SEQ ID NO:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 70, 72, 74, 84, 86, 88, 90, 96, 98, 100, 109, 111, 113, 122, 124, 126, 132, 134, 136, 142, 144, 146, 156, 159, 161, or 163, the cDNA of ATCC® 207192, the cDNA of ATCC® 207223, the cDNA of ATCC® PTA-34, the cDNA of 207189, or the cDNA of ATCC® 207221, or a complement thereof.

In other embodiments, the isolated nucleic acid molecules encode an extracellular, transmembrane, or cytoplasmic domain of a polypeptide of the invention.

In another embodiment, the invention provides an isolated nucleic acid molecule which is antisense to the coding strand of a nucleic acid of the invention.

Another aspect of the invention provides vectors, e.g., recombinant expression vectors, comprising a nucleic acid molecule of the invention. In another embodiment, the invention provides host cells containing such a vector or a nucleic acid molecule of the invention. The invention also provides methods for producing a polypeptide of the invention by culturing, in a suitable medium, a host cell of the invention containing a recombinant expression vector such that a polypeptide is produced.

Another aspect of this invention features isolated or recombinant proteins and polypeptides of the invention. Preferred proteins and polypeptides possess at least one biological activity possessed by the corresponding naturally-occurring human polypeptide. An activity, a biological activity, or a functional activity of a polypeptide or nucleic acid

of the invention refers to an activity exerted by a protein, polypeptide or nucleic acid molecule of the invention on a responsive cell as determined *in vivo*, or *in vitro*, according to standard techniques. Such activities can be a direct activity, such as an association with or an enzymatic activity on a second protein or an indirect activity, such as a cellular signaling activity mediated by interaction of the protein with a second protein.

For TANGO 204, biological activities include, e.g., (1) the ability to form protein-protein interactions with proteins in the signaling pathway of the naturally-occurring polypeptide; (2) the ability to bind a ligand of the naturally-occurring polypeptide; and (3) the ability to interact with a TANGO 204 receptor. Other activities include the ability to modulate function, survival, morphology, proliferation and/or differentiation of cells of tissues in which it is expressed. TANGO 204 biological activities can include the ability to act as a protease inhibitor.

For TANGO 206, biological activities include, e.g., (1) the ability to form protein-protein interactions with proteins in the signaling pathway of the naturally-occurring polypeptide; (2) the ability to bind a ligand of the naturally-occurring polypeptide; and (3) the ability to interact with a TANGO 206 receptor. Other activities include the ability to modulate function, survival, morphology, proliferation and/or differentiation of cells of tissues in which it is expressed. TANGO 206 biological activities can include the ability to modulate cell migration and acid secretion by gastric mucosal tissue.

For TANGO 209, biological activities include, e.g., (1) the ability to form protein-protein interactions with proteins in the signaling pathway of the naturally-occurring polypeptide; (2) the ability to bind a ligand of the naturally-occurring polypeptide; and (3) the ability to interact with a TANGO 209 receptor. Other activities include the ability to modulate function, survival, morphology, proliferation and/or differentiation of cells of tissues in which it is expressed (e.g., cells of the pituitary gland). TANGO 209 biological activities can include the ability to modulate the availability of growth factors, the ability to modulate cell migration, and the ability to modulate embryonic growth.

For A236, biological activities include, e.g., (1) the ability to form protein-protein interactions with proteins in the signaling pathway of the naturally-occurring polypeptide; (2) the ability to bind a ligand of the naturally-occurring polypeptide; and (3)

the ability to interact with a A236 receptor. Other activities include the ability to modulate function, survival, morphology, proliferation and/or differentiation of cells of tissues in which it is expressed. A236 biological activities can include the ability to modulate an inflammatory response and the ability to modulate viral entry.

In one embodiment, a polypeptide of the invention has an amino acid sequence sufficiently identical to an identified domain of a polypeptide of the invention. As used herein, the term "sufficiently identical" refers to a first amino acid or nucleotide sequence which contains a sufficient or minimum number of identical or equivalent (e.g., with a similar side chain) amino acid residues or nucleotides to a second amino acid or nucleotide sequence such that the first and second amino acid or nucleotide sequences have a common structural domain and/or common functional activity. For example, amino acid or nucleotide sequences which contain a common structural domain having about 60% identity, preferably 65% identity, more preferably 75%, 85%, 95%, 98% or more identity are defined herein as sufficiently identical.

In one embodiment, a TANGO 204, TANGO 206, TANGO 209 or A236 polypeptide of the invention includes a signal sequence.

In another embodiment, a nucleic acid molecule of the invention encodes a TANGO 204, TANGO 206, TANGO 209 or A236 polypeptide which includes a signal sequence.

In another embodiment, a TANGO 206 or A236 polypeptide of the invention includes one or more of the following domains: (1) a signal sequence; (2) an extracellular domain; (3) a transmembrane domain; and (4) a cytoplasmic domain.

The polypeptides of the present invention, or biologically active portions thereof, can be operably linked to a heterologous amino acid sequence to form fusion proteins. The invention further features antibodies that specifically bind a polypeptide of the invention such as monoclonal or polyclonal antibodies. In addition, the polypeptides of the invention or biologically active portions thereof, or antibodies of the invention, can be incorporated into pharmaceutical compositions, which optionally include pharmaceutically acceptable carriers.

In another aspect, the present invention provides methods for detecting the presence of the activity or expression of a polypeptide of the invention in a biological

sample by contacting the biological sample with an agent capable of detecting an indicator of activity such that the presence of activity is detected in the biological sample.

In another aspect, the invention provides methods for modulating activity of a polypeptide of the invention comprising contacting a cell with an agent that modulates (inhibits or stimulates) the activity or expression of a polypeptide of the invention such that activity or expression in the cell is modulated. In one embodiment, the agent is an antibody that specifically binds to a polypeptide of the invention.

In another embodiment, the agent modulates expression of a polypeptide of the invention by modulating transcription, splicing, or translation of an mRNA encoding a polypeptide of the invention. In yet another embodiment, the agent is a nucleic acid molecule having a nucleotide sequence that is antisense to the coding strand of an mRNA encoding a polypeptide of the invention.

The present invention also provides methods to treat a subject having a disorder characterized by aberrant activity of a polypeptide of the invention or aberrant expression of a nucleic acid of the invention by administering an agent which is a modulator of the activity of a polypeptide of the invention or a modulator of the expression of a nucleic acid of the invention to the subject. In one embodiment, the modulator is a protein of the invention. In another embodiment, the modulator is a nucleic acid of the invention. In other embodiments, the modulator is a peptide, peptidomimetic, or other small organic molecule.

The present invention also provides diagnostic assays for identifying the presence or absence of a genetic lesion or mutation characterized by at least one of: (i) aberrant modification or mutation of a gene encoding a polypeptide of the invention, (ii) mis-regulation of a gene encoding a polypeptide of the invention, and (iii) aberrant post-translational modification of the invention wherein a wild-type form of the gene encodes a protein having the activity of the polypeptide of the invention.

In another aspect, the invention provides a method for identifying a compound that binds to or modulates the activity of a polypeptide of the invention. In general, such methods entail measuring a biological activity of the polypeptide in the presence and absence of a test compound and identifying those compounds which alter the activity of the polypeptide.

The invention also features methods for identifying a compound which modulates the expression of a polypeptide or nucleic acid of the invention by measuring the expression of the polypeptide or nucleic acid in the presence and absence of the compound.

In yet a further aspect, the invention provides substantially purified antibodies or fragments thereof, including human and non-human antibodies or fragments thereof, which antibodies or fragments specifically bind to a polypeptide comprising an amino acid sequence selected from the group consisting of: the amino acid sequence of SEQ ID Nos. 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 12, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164, or the amino acid sequence encoded by the cDNA of ATCC® 207192, the amino acid sequence encoded by the cDNA of ATCC® 207223, the amino acid sequence encoded by the cDNA of ATCC® PTA-34, the amino acid sequence encoded by the cDNA of ATCC® 207189, or the amino acid sequence encoded by the cDNA of ATCC® 207221; a fragment of at least 15 amino acid residues of the amino acid sequence of SEQ ID Nos. 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 12, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164; an amino acid sequence which is at least 95% identical to the amino acid sequence of SEQ ID Nos. 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 12, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164, wherein the percent identity is determined using the ALIGN program of the GCG software package with a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4; and an amino acid sequence which is encoded by a nucleic acid molecule which hybridizes to the nucleic acid molecule consisting of SEQ ID NO:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 70, 72, 74, 84, 86, 88, 90, 96, 98, 100, 109, 111, 113, 122, 124, 126, 132, 134, 136, 142, 144, 146, 156, 159, 161, or 163, under conditions of hybridization of 6X SSC at 45°C and washing in 0.2 X SSC, 0.1% SDS at 65°C. In various embodiments, the substantially purified antibodies of the invention, or fragments thereof, can be human, non-human, chimeric and/or humanized antibodies.

In another aspect, the invention provides non-human antibodies or fragments thereof, which antibodies or fragments specifically bind to a polypeptide comprising an amino acid sequence selected from the group consisting of: the amino acid sequence of SEQ ID Nos. 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 12, 114, 123,

125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164, or the amino acid sequence encoded by the cDNA of ATCC® 207192, the amino acid sequence encoded by the cDNA of ATCC® 207223, the amino acid sequence encoded by the cDNA of ATCC® PTA-34, the amino acid sequence encoded by the cDNA of ATCC® 207189, or the amino acid sequence encoded by the cDNA of ATCC® 207221; a fragment of at least 15 amino acid residues of the amino acid sequence of SEQ ID Nos. 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 12, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164; an amino acid sequence which is at least 95% identical to the amino acid sequence of SEQ ID Nos. 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 12, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164, wherein the percent identity is determined using the ALIGN program of the GCG software package with a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4; and an amino acid sequence which is encoded by a nucleic acid molecule which hybridizes to the nucleic acid molecule consisting of SEQ ID NO:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 70, 72, 74, 84, 86, 88, 90, 96, 98, 100, 109, 111, 113, 122, 124, 126, 132, 134, 136, 142, 144, 146, 156, 159, 161, or 163, under conditions of hybridization of 6X SSC at 45°C and washing in 0.2 X SSC, 0.1% SDS at 65°C. Such non-human antibodies can be goat, mouse, sheep, horse, chicken, rabbit, or rat antibodies. Alternatively, the non-human antibodies of the invention can be chimeric and/or humanized antibodies. In addition, the non-human antibodies of the invention can be polyclonal antibodies or monoclonal antibodies.

In still a further aspect, the invention provides monoclonal antibodies or fragments thereof, which antibodies or fragments specifically bind to a polypeptide comprising an amino acid sequence selected from the group consisting of: the amino acid sequence of SEQ ID Nos. 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 12, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164, or the amino acid sequence encoded by the cDNA of ATCC® 207192, the amino acid sequence encoded by the cDNA of ATCC® 207223, the amino acid sequence encoded by the cDNA of ATCC® PTA-34, the amino acid sequence encoded by the cDNA of ATCC® 207189, or the amino acid sequence encoded by the cDNA of ATCC® 207221; a fragment of at least 15 amino acid residues of the amino acid sequence of SEQ ID Nos. 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 12, 114, 123, 125, 127, 133, 135, 137, 143, 145,

147, 157, 160, 162, or 164; an amino acid sequence which is at least 95% identical to the amino acid sequence of SEQ ID Nos. 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 12, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164, wherein the percent identity is determined using the ALIGN program of the GCG software package with a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4; and an amino acid sequence which is encoded by a nucleic acid molecule which hybridizes to the nucleic acid molecule consisting of SEQ ID NO:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 70, 72, 74, 84, 86, 88, 90, 96, 98, 100, 109, 111, 113, 122, 124, 126, 132, 134, 136, 142, 144, 146, 156, 159, 161, or 163, under conditions of hybridization of 6X SSC at 45°C and washing in 0.2 X SSC, 0.1% SDS at 65°C. The monoclonal antibodies can be human, humanized, chimeric and/or non-human antibodies.

In a particularly preferred embodiment, the substantially purified antibodies or fragments thereof, the human or non-human antibodies or fragments thereof, and/or the monoclonal antibodies or fragments thereof, of the invention specifically bind to an extracellular domain of the amino acid sequence of SEQ ID NO:42, 46, 59, 65, 68, or 171. Preferably, the extracellular domain to which the antibody, or fragment thereof, binds comprises amino acid residues 30-362 of SEQ ID NO:42, amino acid residues 406-420 of SEQ ID NO:46, amino acid residues 19-230 of SEQ ID NO:59, amino acid residues 30-362 of SEQ ID NO:65; amino acid residues 406-420 of SEQ ID NO:68, or amino acid residues 18-229 of SEQ ID NO:171.

Any of the antibodies of the invention can be conjugated to a therapeutic moiety or to a detectable substance. Non-limiting examples of detectable substances that can be conjugated to the antibodies of the invention are an enzyme, a prosthetic group, a fluorescent material, a luminescent material, a bioluminescent material, and a radioactive material.

The invention also provides a kit containing an antibody of the invention conjugated to a detectable substance, and instructions for use. Still another aspect of the invention is a pharmaceutical composition comprising an antibody of the invention and a pharmaceutically acceptable carrier. In one embodiment, the pharmaceutical composition contains an antibody of the invention, a therapeutic moiety, and a pharmaceutically acceptable carrier.

In addition, the polypeptides of the invention or biologically active portions thereof, or antibodies of the invention, can be incorporated into pharmaceutical compositions, which optionally include pharmaceutically acceptable carriers.

The invention also features transgenic non-human animals which include (and preferably express) a heterologous form of a TANGO 204, TANGO 206, TANGO 209, or A236 gene described herein, or which misexpress (*e.g.*, do not express) an endogenous TANGO 204, TANGO 206, TANGO 209, or A236 gene (*e.g.*, an animal in which expression of one or more of the subject TANGO 204, TANGO 206, TANGO 209, or A236 proteins is disrupted). Such a transgenic animal can serve as an animal model for studying cellular or tissue disorders comprising mutated or mis-expressed TANGO 204, TANGO 206, TANGO 209, or A236 alleles or for use in drug screening. For example, the transgenic animals of the invention can be used as an animal model to study a TANGO 204, TANGO 206, TANGO 209, or A236-mediated disease. Alternatively, such a transgenic animal can be useful for expressing recombinant polypeptides, and for generating cells, *e.g.*, cell lines that can, for example, be utilized as part of screening techniques.

Specifically, for example, for TANGO 204 family members, diseases which may be studied include, without limitation, those associated with tissues of the eye, lung, stomach, intestine and developing bone and cartilage structures such as the ear, nose, and spinal column. For TANGO 206 family members, diseases which may be studied include, without limitation, those associated with tissues of the heart, skeletal muscle, kidney, brain, and placenta, central nervous system, eye, large ganglion of the head, liver and skin. For TANGO 209 family members, diseases which may be studied include, without limitation, those associated with tissues of the heart, skeletal muscle, pancreas, placenta, lung, kidney, brain (hippocampus, dentate gyrus, and frontal cortex), thymus (multifocal expression), kidney (medulla and capsule), adrenal gland (capsule), spleen (non-follicular), bladder, small intestine and colon (smooth muscle, not villi), liver. For A236 family members, diseases which may be studied include, without limitation, those associated with tissues of the brain, placenta, uterus, ovary, mandible, and the intestinal tract.

In another aspect, the invention provides a method for identifying a compound that binds to or modulates the activity of a polypeptide of the invention. In general, such

methods entail measuring a biological activity of the polypeptide in the presence and absence of a test compound and identifying those compounds which alter the activity of the polypeptide.

The invention also features methods for identifying a compound which modulates the expression of a polypeptide or nucleic acid of the invention by measuring the expression of the polypeptide or nucleic acid in the presence and absence of the compound.

In one embodiment, the invention provides assays, *e.g.*, for screening test compounds to identify agonists, or alternatively, antagonists, of a TANGO 204, TANGO 206, TANGO 209, or A236 activity. For example, the test compound may positively or negatively influence an interaction between a TANGO 204, TANGO 206, TANGO 209, or A236 protein and a TANGO 204, TANGO 206, TANGO 209, or A236 target molecule. An exemplary method includes the steps of (i) combining a TANGO 204, TANGO 206, TANGO 209, or A236 polypeptide or active fragment thereof, with a TANGO 204, TANGO 206, TANGO 209, or A236 target molecule, and a test compound, *e.g.*, under conditions wherein, but for the test compound, the TANGO 204, TANGO 206, TANGO 209, or A236 protein and target molecule are able to interact; and (ii) detecting the formation of a complex which includes the TANGO 204, TANGO 206, TANGO 209, or A236 protein and the target molecule either by directly quantitating the complex, by measuring inductive effects of the TANGO 204, TANGO 206, TANGO 209, or A236 protein, or, in the instance of a substrate, measuring the conversion to product. A statistically significant change, such as a decrease, in the interaction of the TANGO 204, TANGO 206, TANGO 209, or A236 and target molecule in the presence of a test compound (relative to what is detected in the absence of the test compound) is indicative of a modulation (*e.g.*, suppression or potentiation of the interaction between the TANGO 204, TANGO 206, TANGO 209, or A236 protein and the target molecule).

The invention provides yet other methods for identifying compounds which modulate a TANGO 204, TANGO 206, TANGO 209, or A236 activity. For example, a compound that interacts with a TANGO 204, TANGO 206, TANGO 209, or A236 protein can be identified by contacting a TANGO 204, TANGO 206, TANGO 209, or A236 protein with a test compound. Either the test compound or the TANGO 204, TANGO 206, TANGO 209, or A236 protein can be labeled. Optionally, the non-labeled

compound or TANGO 204, TANGO 206, TANGO 209, or A236 protein can be attached to a solid phase support. Binding of the test compound to the TANGO 204, TANGO 206, TANGO 209, or A236 protein can then be determined, e.g by measuring the amount of label. Such a TANGO 204, TANGO 206, TANGO 209, or A236 binding molecule can be an agonist or an antagonist. In one embodiment, an agonist of a TANGO 204, TANGO 206, TANGO 209, or A236 activity is identified by contacting a cell having a TANGO 204, TANGO 206, TANGO 209, or A236 protein with a test compound and measuring a TANGO 204, TANGO 206, TANGO 209, or A236 activity, *e.g.*, expression of a gene which is regulated by binding of a protein to TANGO 204, TANGO 206, TANGO 209, or A236. An increased expression of the gene when the cell is incubated with the test compound relative to incubation in the absence of the test compound indicates that the test compound is a TANGO 204, TANGO 206, TANGO 209, or A236 agonist. The gene that is monitored can also be a reporter gene transfected to a cell, the reporter gene being under the control of a promoter of a gene which is regulated by TANGO 204, TANGO 206, TANGO 209, or A236.

In another aspect, the invention provides methods for modulating activity of a polypeptide of the invention comprising contacting a cell with an agent that modulates (inhibits or stimulates) the activity or expression of a polypeptide of the invention such that activity or expression in the cell is modulated. In one embodiment, the agent is an antibody that specifically binds to a polypeptide of the invention.

In another embodiment, the agent modulates expression of a polypeptide of the invention by modulating transcription, splicing, or translation of an mRNA encoding a polypeptide of the invention. In yet another embodiment, the agent is a nucleic acid molecule having a nucleotide sequence that is antisense to the coding strand of an mRNA encoding a polypeptide of the invention.

Other features and advantages of the invention will be apparent from the following detailed description and claims.

Brief Description of the Drawings

Figures 1A-1B depict the cDNA sequence (SEQ ID NO:1) and the predicted amino acid sequence (SEQ ID NO:2) of human TANGO 204. The open reading frame of

SEQ ID NO:1 extends from nucleotide 99 to nucleotide 890 of SEQ ID NO:1 (SEQ ID NO:3).

Figure 2 depicts a hydropathy plot of human TANGO 204. Relatively hydrophobic regions are above the dashed horizontal line, and relatively hydrophilic regions are below the dashed horizontal line. The cysteine residues (cys) are indicated by short vertical lines just below the hydropathy trace.

Figure 3 depicts an alignment of the somatomedin B domain of human TANGO 204 (SEQ ID NO:36) with a consensus somatomedin B domain (SEQ ID NO:37). In the consensus sequence, more conserved residues are indicated by uppercase letters, and less conserved residues are indicated by lowercase letters.

Figure 4 depicts an alignment of the thrombospondin type 1 domain of human TANGO 204 (SEQ ID NO:38) with a consensus thrombospondin type 1 domain (SEQ ID NO:39). In the consensus sequence, more conserved residues are indicated by uppercase letters, and less conserved residues are indicated by lowercase letters.

Figures 5A-5B depict the cDNA sequence (SEQ ID NO:13) and the predicted amino acid sequence (SEQ ID NO:14) of murine TANGO 204. The open reading frame of SEQ ID NO:13 extends from nucleotide 81-872 of SEQ ID NO:13 (SEQ ID NO:15).

Figures 6A-6B depict an alignment of the open reading frames of human TANGO 204 (SEQ ID NO:3) and murine TANGO 204 (SEQ ID NO:15).

Figure 7 depicts an alignment of the amino acid sequences of human TANGO 204 (SEQ ID NO:2) and murine TANGO 204 (SEQ ID NO:14).

Figures 8A-8B depict the cDNA sequence (SEQ ID NO:4) and the predicted amino acid sequence (SEQ ID NO:5) of human TANGO 206. The open reading frame of SEQ ID NO:4 extends from nucleotide 99 to nucleotide 1358 of SEQ ID NO:4 (SEQ ID NO:6).

Figure 9 depicts a hydropathy plot of human TANGO 206. Relatively hydrophobic regions are above the dashed horizontal line, and relatively hydrophilic regions are below the dashed horizontal line. The cysteine residues (cys) and potential N-glycosylation sites (Ngly) are indicated by short vertical lines just below the hydropathy trace.

Figure 10 depicts an alignment of the laminin EGF-like domain of human TANGO 206 (SEQ ID NO:47) with a consensus laminin EGF-like domain (SEQ ID

NO:48). In the consensus sequence, more conserved residues are indicated by uppercase letters, and less conserved residues are indicated by lowercase letters.

Figures 11A-11B depict the cDNA sequence (SEQ ID NO:16) and the predicted amino acid sequence (SEQ ID NO:17) of murine TANGO 206. The open reading frame of SEQ ID NO:16 extends from nucleotide 332-1591 of SEQ ID NO:16 (SEQ ID NO:18).

Figures 12A-12C depict an alignment of the open reading frames of human TANGO 206 (SEQ ID NO:6) and murine TANGO 206 (SEQ ID NO:18).

Figure 13 depicts an alignment of the amino acid sequences of human TANGO 206 (SEQ ID NO:5) and murine TANGO 206 (SEQ ID NO:17).

Figures 14A-14C depict the cDNA sequence (SEQ ID NO:7) and the predicted amino acid sequence (SEQ ID NO:8) of human TANGO 209. The open reading frame of SEQ ID NO:7 extends from nucleotide 194-1531 of SEQ ID NO:7 (SEQ ID NO:9).

Figure 15 depicts a hydropathy plot of human TANGO 209. Relatively hydrophobic regions are above the dashed horizontal line, and relatively hydrophilic regions are below the dashed horizontal line. The cysteine residues (cys) and potential N-glycosylation sites (Ngly) are indicated by short vertical lines just below the hydropathy trace.

Figure 16 depicts an alignment of the thyroglobulin type 1 domains of human TANGO 209 (SEQ ID NO:51 and SEQ ID NO:52) with a consensus thyroglobulin type 1 domain (SEQ ID NO:53). In the consensus sequence, more conserved residues are indicated by uppercase letters, and less conserved residues are indicated by lowercase letters.

Figure 17 depicts an alignment of the Kazal-type serine protease inhibitor domains of human TANGO 209 (SEQ ID NO:54) with a consensus Kazal-type serine protease inhibitor domain (SEQ ID NO:55). In the consensus sequence, more conserved residues are indicated by uppercase letters, and less conserved residues are indicated by lowercase letters.

Figures 18A-18C depict the cDNA sequence (SEQ ID NO:19) and the predicted amino acid sequence (SEQ ID NO:20) of murine TANGO 209. The open reading frame of SEQ ID NO:19 extends from nucleotide 187 to nucleotide 1527 of SEQ ID NO:19 (SEQ ID NO:21).

Figures 19A-19C depict an alignment of the open reading frames of human TANGO 209 (SEQ ID NO:9) and murine TANGO 209 (SEQ ID NO:21).

Figure 20 depicts an alignment of the amino acid sequences of human TANGO 209 (SEQ ID NO:8) and murine TANGO 209 (SEQ ID NO:20).

Figures 21A-21B depict the cDNA sequence (SEQ ID NO:10) and the predicted amino acid sequence (SEQ ID NO:11) of human A236. The open reading frame of SEQ ID NO:10 extends from nucleotide 314 to nucleotide 1432 of SEQ ID NO:10 (SEQ ID NO:12).

Figure 22 depicts a hydropathy plot of human A236. Relatively hydrophobic regions are above the dashed horizontal line, and relatively hydrophilic regions are below the dashed horizontal line. The cysteine residues (cys) and potential N-glycosylation sites (Ngly) are indicated by short vertical lines just below the hydropathy trace.

Figure 23 depicts an alignment of the immunoglobulin domains of human A236 (SEQ ID NO:62 and SEQ ID NO:63) with a consensus immunoglobulin domain (SEQ ID NO:64).

Figures 24A-24B depict the cDNA sequence (SEQ ID NO:156) and the predicted amino acid sequence (SEQ ID NO:157) of mouse A236. The open reading frame of SEQ ID NO:156 extends from nucleotide to nucleotide 304-1422 of SEQ ID NO:156 (SEQ ID NO:158).

*Figures 25A-25*B depict an alignment of the open reading frames of human A236 (SEQ ID NO:12) and murine A236 (SEQ ID NO:158).

Figure 26 depicts an alignment of the amino acid sequences of human A236 (SEQ ID NO:11) and murine A236 (SEQ ID NO:157).

Figure 27 depicts a hydropathy plot of mouse A236. Relatively hydrophobic regions are above the dashed horizontal line, and relatively hydrophilic regions are below the dashed horizontal line. The cysteine residues (cys) and potential N-glycosylation sites (Ngly) are indicated by short vertical lines just below the hydropathy trace.

Definitions

For convenience, the meaning of certain terms and phrases employed in the specification, examples, and appended claims are provided below.

The term "allele", which is used interchangeably herein with "allelic variant" refers to alternative forms of a gene, nucleic acid or portions thereof, as well as to a polypeptide encoded by said gene, nucleic acid, or portion thereof. Nucleic acid alleles occupy the same locus or position on homologous chromosomes. When a subject has two identical alleles of a gene, the subject is said to be homozygous for the gene or allele. When a subject has two different alleles of a gene, the subject is said to be heterozygous for the gene. Alleles of a specific gene can differ from each other in a single nucleotide, or several nucleotides, and can include substitutions, deletions, and insertions of nucleotides. An allele of a gene can also be a form of a gene containing a mutation.

The term "allelic variant of a polymorphic region of a TANGO 204, TANGO 206, TANGO 209, or A236 gene" refers to a region of a TANGO 204, TANGO 206, TANGO 209, or A236 gene having one of several nucleotide sequences found in that region of the gene in other individuals, as well as to polypeptides encoded by nucleic acid molecules comprising said sequences.

The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, *i.e.*, molecules that contain an antigen binding site which specifically binds an antigen, such as a polypeptide of the invention, *e.g.*, an epitope of a polypeptide of the invention. A molecule which specifically binds to a given polypeptide of the invention is a molecule which binds the polypeptide, but does not substantially bind other molecules in a sample, *e.g.*, a biological sample, which naturally contains the polypeptide. Examples of immunologically active portions of immunoglobulin molecules include F(ab) and $F(ab')_2$ fragments which can be generated by treating the antibody with an enzyme such as pepsin. The invention provides polyclonal and monoclonal antibodies. The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope.

"Cells," "host cells" or "recombinant host cells" are terms used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such

progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

"Complementary" sequences as used herein refer to sequences which have sufficient complementarity to be able to hybridize, forming a stable duplex.

A "delivery complex" shall mean a targeting means (*e.g.*, a molecule that results in higher affinity binding of a gene, protein, polypeptide or peptide to a target cell surface and/or increased cellular uptake by a target cell). Examples of targeting means include: sterols (*e.g.*, cholesterol), lipids (*e.g.*, a cationic lipid, virosome or liposome), viruses (*e.g.*, adenovirus, adeno-associated virus, and retrovirus) or target cell specific binding agents (*e.g.*, ligands recognized by target cell specific receptors). Preferred complexes are sufficiently stable *in vivo* to prevent significant uncoupling prior to internalization by the target cell. However, the complex is cleavable under appropriate conditions within the cell so that the gene, protein, polypeptide or peptide is released in a functional form. It is also possible that soluble forms of the protein also exist. Such soluble isoforms can arise through variable splicing of the TANGO 204, TANGO 206, TANGO 209, or A236 gene or alternatively as a result of proteolysis of a membranous isoform.

As is well known, genes for a particular polypeptide may exist in single or multiple copies within the genome of an individual. Such duplicate genes may be identical or may have certain modifications, including nucleotide substitutions, additions or deletions, which all still code for polypeptides having substantially the same activity. The term "DNA sequence encoding a TANGO 204, TANGO 206, TANGO 209, or A236 polypeptide" may thus refer to one or more genes within a particular individual. Moreover, certain differences in nucleotide sequences may exist between individual organisms, which are called alleles. Such allelic differences may or may not result in differences in amino acid sequence of the encoded polypeptide yet still encode a protein with the same biological activity.

As used herein, the term "gene" or "recombinant gene", as applied to TANGO 204, TANGO 206, TANGO 209, or A236, refers to a polynucleotide or nucleic acid molecule comprising an open reading frame encoding one of the TANGO 204, TANGO 206, TANGO 209, or A236 polypeptides of the present invention. In one embodiment, these terms relate to a cDNA sequence including, but not limited to, a polynucleotide or

nucleic acid sequence obtained via reverse transcription of an mRNA molecule. In one embodiment, the term nucleic acid or polynucleotide is a nucleic acid molecule which is not genomic but is a cDNA derived from a contiguous coding region which includes, but is not limited to, reverse transcribed cDNA. In another embodiment, the term nucleic acid or polynucleotide refers to a nucleic acid molecule which comprises contiguous nucleotide codons. In yet another embodiment, the term nucleic acid or polynucleotide is a nucleic acid molecule which is genomic but which excludes intronic sequences.

"Homology" or "identity" or "similarity" refers to sequence similarity between two peptides or between two nucleic acid molecules. Homology can be determined by comparing a position in each sequence which may be aligned for purposes of comparison. When a position in the compared sequence is occupied by the same base or amino acid, then the molecules are identical at that position. A degree of homology or similarity or identity between nucleic acid sequences is a function of the number of identical or matching nucleotides at positions shared by the nucleic acid sequences. A degree of identity of amino acid sequences is a function of the number of identical amino acids at positions shared by the amino acid sequences. Likewise, a degree of identity of nucleic acid sequences is a function of the number of identical nucleic acids at positions shared by the nucleic acid sequences.

Furthermore, a degree of homology or similarity of amino acid sequences is a function of the number of conserved amino acids at positions shared by the amino acid sequences. A sequence which is "unrelated" or "non-homologous" with one of the human TANGO 204, TANGO 206, TANGO 209, or A236 sequences of the present invention typically is a sequence which shares less than 40 % identity, though preferably less than 25 % identity with one of the human TANGO 204, TANGO 206, TANGO 209, or A236 sequences of the present invention.

To determine the percent identity of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the

molecules are identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences (*i.e.*, % identity = # of identical positions/total # of positions (*e.g.*, overlapping positions) x 100).

Preferably, the determination of percent identity between two sequences is accomplished using a mathematical algorithm. A preferred, non-limiting example of a mathematical algorithm utilized for the comparison of two sequences is the algorithm of Karlin and Altschul (1990) *Proc. Natl. Acad. Sci. USA* 87:2264-2268, modified as in Karlin and Altschul (1993) *Proc. Natl. Acad. Sci. USA* 90:5873-5877. Such an algorithm is incorporated into the NBLAST and XBLAST programs of Altschul, et al. (1990) *J. Mol. Biol.* 215:403-410. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to a nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to a protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al. (1997) *Nucleic Acids Res.* 25:3389-3402. Alternatively, PSI-Blast can be used to perform an iterated search which detects distant relationships between molecules. *Id.* When utilizing BLAST, Gapped BLAST, and PSI-Blast programs, the default parameters of the respective programs (*e.g.*, XBLAST and NBLAST) can be used. *See* http://www.ncbi.nlm.nih.gov. Another preferred, non-limiting example of a mathematical algorithm utilized for the comparison of sequences is the algorithm of Myers and Miller, (1988) *CABIOS* 4:11-17. Such an algorithm is incorporated into the ALIGN program (version 2.0) which is part of the GCG sequence alignment software package. When utilizing the ALIGN program for comparing amino acid sequences, a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 can be used.

The percent identity between two sequences can be determined using techniques similar to those described above, with or without allowing gaps. In calculating percent identity, typically only exact matches are counted.

As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least 60% (65%, 70%, preferably 75% or more) identical to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art

and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. A preferred, non-limiting example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45°C, followed by one or more washes in 2.0 X SSC at 50° C. (low stringency) or 0.2 X SSC, 0.1% SDS at 50-65°C (high stringency). In one embodiment, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequence of SEQ ID Nos. 1, 3, 4, 6, 7; 9, 10, 12, 13, 15, 16, 18, 19, 21, or 156, or complement thereof, corresponds to a naturally-occurring nucleic acid molecule.

Detailed Description of the Invention

The present invention is based, at least in part, on the discovery of cDNA molecules encoding TANGO 204, TANGO 206, TANGO 209, and A236, all of which are predicted to be either wholly secreted or transmembrane proteins.

The proteins and nucleic acid molecules of the present invention comprise a family of molecules having certain conserved structural and functional features. As used herein, the term "family" is intended to mean two or more proteins or nucleic acid molecules having a common structural domain and having sufficient amino acid or nucleotide sequence identity as defined herein. Family members can be from either the same or different species. For example, a family can comprise two or more proteins of human origin, or can comprise one or more proteins of human origin and one or more of non-human origin.

Members of the same family may also have common structural domains. For example, a TANGO 204, TANGO 206, TANGO 209, or A236 family member includes a signal sequence. As used herein, a "signal sequence" includes a peptide of at least about 15 amino acid residues in length which occurs at the N-terminus of secretory and membrane-bound proteins and which contains at least about 70% hydrophobic amino acid residues such as alanine, leucine, isoleucine, phenylalanine, proline, tyrosine, tryptophan, or valine. The sequence can contain about 15 to 30 amino acid residues or about 17-22 amino acid residues, and has at least about 60-80%, 65-75%, or about 70% hydrophobic residues. A signal sequence serves to direct a protein containing such a sequence to a lipid bilayer.

Thus, in one embodiment, a TANGO 204 protein contains a signal sequence of about amino acids 1-20 of SEQ ID NO:2 (SEQ ID NO:22). In certain embodiments, a TANGO 204 family member has the amino acid sequence of SEQ ID NO:2, and the signal sequence is located at amino acids 1 to 18, 1 to 19, 1 to 20, 1 to 21 or 1 to 22. In such embodiments of the invention, the domains and the mature protein resulting from cleavage of such signal peptides are also included herein. For example, the cleavage of a signal sequence consisting of amino acids 1 to 20 of SEQ ID NO:2 (SEQ ID NO:22) results in a mature TANGO 204 protein corresponding to amino acids 21 to 264 of SEQ ID NO:2 (SEQ ID NO:23). The signal sequence is normally cleaved during processing of the mature protein.

In one embodiment, a TANGO 206 protein contains a signal sequence of about amino acids 1-29 of SEQ ID NO:5 (SEQ ID NO:24). In certain embodiments, a TANGO 206 family member has the amino acid sequence of SEQ ID NO:5, and the signal sequence is located at amino acids 1 to 27, 1 to 28, 1 to 29, 1 to 30 or 1 to 31. In such embodiments of the invention, the domains and the mature protein resulting from cleavage of such signal peptides are also included herein. For example, the cleavage of a signal sequence consisting of amino acids 1 to 29 of SEQ ID NO:5 (SEQ ID NO:24) results in a mature TANGO 206 protein corresponding to amino acids 30 to 420 of SEQ ID NO:5 (SEQ ID NO:25). The signal sequence is normally cleaved during processing of the mature protein.

In one embodiment, a TANGO 209 protein contains a signal sequence of about amino acids 1-21 of SEQ ID NO:8 (SEQ ID NO:26). In certain embodiments, a TANGO 209 family member has the amino acid sequence of SEQ ID NO:8, and the signal sequence is located at amino acids 1 to 19, 1 to 20, 1 to 21, 1 to 22 or 1 to 23. In such embodiments of the invention, the domains and the mature protein resulting from cleavage of such signal peptides are also included herein. For example, the cleavage of a signal sequence consisting of amino acids 1 to 21 of SEQ ID NO:8 (SEQ ID NO:26) results in a mature TANGO 209 protein corresponding to amino acids 22 to 446 of SEQ ID NO:8 (SEQ ID NO:27). The signal sequence is normally cleaved during processing of the mature protein.

In one embodiment, a A236 protein contains a signal sequence of amino acids 1-18 of SEQ ID NO:11 (SEQ ID NO:28). The signal sequence is cleaved during

processing of the mature protein. In certain embodiments, an A236 family member has the amino acid sequence of SEQ ID NO:11, and the signal sequence is located at amino acids 1 to 16, 1 to 17, 1 to 18, 1 to 19 or 1 to 20. In such embodiments of the invention, the domains and the mature protein resulting from cleavage of such signal peptides are also included herein. For example, the cleavage of a signal sequence consisting of amino acids 1 to 18 of SEQ ID NO:11 (SEQ ID NO:28) results in a mature A236 protein corresponding to amino acids 19 to 373 of SEQ ID NO:11 (SEQ ID NO:29). The signal sequence is normally cleaved during processing of the mature protein.

In another example, a TANGO 206 or A236 family member also includes one or more of the following domains: (1) an extracellular domain; (2) a transmembrane domain; and (3) a cytoplasmic domain.

In one embodiment, a TANGO 206 protein contains an extracellular domain of about amino acids 30-362 of SEQ ID NO:5 (SEQ ID NO:42). In one embodiment, a TANGO 206 protein contains a transmembrane of about amino acids 363-379 of SEQ ID NO:5 (SEQ ID NO:43). In another embodiment, a TANGO 206 protein contains a cytoplasmic domain of about amino acids 380-386 of SEQ ID NO:5 (SEQ ID NO:44). In another embodiment, a TANGO 206 protein includes a transmembrane domain of about amino acids 387-405 of SEQ ID NO:5 (SEQ ID NO:45). In still another embodiment, a TANGO 206 protein includes an extracellular domain of about amino acids 406-420 of SEQ ID NO:5 (SEQ ID NO:46).

In one embodiment, a A236 protein includes an extracellular domain of about amino acids 19-230 of SEQ ID NO:11 (SEQ ID NO:59). In another embodiment, a TANGO 206 protein includes a transmembrane domain of about amino acids 231-255 of SEQ ID NO:11 (SEQ ID NO:60). In yet another embodiment, a A236 protein includes a cytoplasmic domain of about amino acids 256-373 of SEQ ID NO:11 (SEQ ID NO:61).

TANGO 204 family members can also include a somatomedin B domain. Somatomedin B domains are present in plasma cell glycoprotein PC-1 and placental protein 11. Somatomedin B domains have the sequence Cys-Xaa_6-C-Xaa_9-Cys-Xaa-Cys-Xaa_3-Cys-Xaa_5-Cys-Cys-Xaa_5-Cys (where Xaa can be any amino acid). The most highly conserved portion of the somatomedin B domain has the sequence Cys-Xaa-Cys-Xaa_3-C-Xaa_4-Cys-Cys-Xaa_4-Cys (where Xaa can be any amino acid). The cysteine residues within the domain are all likely involved in disulfide bonds. A consensus somatomedin B

domain has the sequence of SEQ ID NO:37. This consensus sequence is shown in Figure 3 where the more conserved residues in the consensus sequence are indicated by uppercase letters and the less conserved residues in the consensus sequence are indicated by lowercase letters. The somatomedin B domain of human TANGO 204 is located at amino acids 18-75 of SEQ ID NO:2 (SEQ ID NO:36).

TANGO 204 family members can also include a thrombospondin type I domain. A consensus thrombospondin type 1 domain has the sequence of SEQ ID NO:39. This consensus sequence is shown in Figure 4 where the more conserved residues in the consensus sequence are indicated by uppercase letters and the less conserved residues inthe consensus sequence are indicated by lowercase letters. The thrombospondin type 1 domain of human TANGO 204 is located at amino acids 78-121 of SEQ ID NO:2 (SEQ ID NO:38). Thrombospondin type 1 domains can include the sequence CS(A/V)TCG and the sequence W(S/G)XW.

TANGO 206 family members can include a laminin EGF-like domain. A consensus laminin EGF-like domain has the sequence of SEQ ID NO:48. This consensus sequence is shown in Figure 10 where the more conserved residues in the consensus sequence are indicated by uppercase letters and the less conserved residues inthe consensus sequence are indicated by lowercase letters. The laminin EGF-like domain of human TANGO 204 is located at amino acids 168-211 of SEQ ID NO:5 (SEQ ID NO:47). Laminin EGF-like domains are similar to EGF domains except that they include eight cysteines rather than 6 cysteines. All eight cysteines are expected to participate in disulfide bonds.

TANGO 209 family members can include a thyroglobulin type 1 domain. A consensus thyroglobulin type 1 domain has the sequence of SEQ ID NO:53. This consensus sequence is shown in Figure 16 where the more conserved residues in the consensus sequence are indicated by uppercase letters and the less conserved residues in the consensus sequence are indicated by lowercase letters. The thyroglobulin type 1 domains of TANGO 209 are located at amino acids 109-153 of SEQ ID NO:8 (SEQ ID NO:51) and amino acids 237-281 of SEQ ID NO:8 (SEQ ID NO:52). The thyroglobulin type 1 domain is present in HLA class II associate invariant chain, HLA class II associated invariant chain, and pancreatic carcinoma marker proteins GA733-1 and GA733-2.

TANGO 209 family members can include a Kazal-type serine protease inhibitor domain. A consensus Kazal-type serine protease inhibitor domain has the sequence of SEQ ID NO:55. This consensus sequence is shown in Figure 17 where the more conserved residues in the consensus sequence are indicated by uppercase letters and the less conserved residues inthe consensus sequence are indicated by lowercase letters. The Kazal-type serine protease inhibitor domain of TANGO 209 is located at amino acids 40-84 of SEQ ID NO:8 (SEQ ID NO:54).

A236 family members can include a immunoglobulin domain. Immunoglobulin domains are present in a variety of proteins and are involved in protein-protein and protein-ligand interaction. A consensus immunoglobulin domain has the sequence of SEQ ID NO:64. This consensus sequence is shown in Figure 23 where the more conserved residues in the consensus sequence are indicated by uppercase letters and the less conserved residues in the consensus sequence are indicated by lowercase letters. The immunoglobulin domains of human A236 are located at amino acids 28-113 of SEQ ID NO:11 (SEQ ID NO:62) and amino acids 146-210 of SEQ ID NO:11 (SEQ ID NO:63). The immunoglobulin domains of mouse A236 are located at amino acids 27-112 of SEQ ID NO:157 (SEQ ID NO:176) and amino acids 145-209 of SEQ ID NO:157 (SEQ ID NO:177).

Various features of TANGO 204, TANGO 206, TANGO 209, and A236 are summarized below.

Human TANGO 204

A cDNA encoding TANGO 204 was identified by analyzing the sequences of clones present in a human lung cDNA library.

This analysis led to the identification of a clone, Athu204c, encoding full-length human TANGO 204. The cDNA of this clone is 3057 nucleotides long (Figures 1A-1B; SEQ ID NO:1). The 792 nucleotide open reading frame of this cDNA, nucleotides 99-890 of SEQ ID NO:1 (SEQ ID NO:3), encodes a 264 amino acid protein (Figures 1A-1B; SEQ ID NO:2).

In one embodiment of a nucleotide sequence of human TANGO 204 the nucleotide at position 170 is a guanine (G)(SEQ ID NO:1). In this embodiment, the amino acid at position 24 is glutamate (E)(SEQ ID NO:2). In another embodiment of a

nucleotide sequence of human TANGO 204 the nucleotide at position 170 is a cytosine (C)(SEQ ID NO:70). In this embodiment, the amino acid at position 24 is aspartate (D)(SEQ ID NO:71) In another embodiment of a nucleotide sequence of human TANGO 204, the nucleotide at position 335 is an adenine (A)(SEQ ID NO:1). In this embodiment, the amino acid at position 79 is a glutamate (E)(SEQ ID NO:2). In another embodiment of a nucleotide sequence of human TANGO 204, the nucleotide at position 335 is a cytosine (C)(SEQ ID NO:72). In this embodiment, the amino acid at position 79 is aspartate (D)(SEQ ID NO:73). In another embodiment of a nucleotide sequence of human TANGO 204, the nucleotide at position 410 is a guanine (G)(SEQ ID NO:1). In this embodiment, the amino acid at position 104 is a glutamate (E)(SEQ ID NO:2). In another embodiment of a nucleotide sequence of human TANGO 204, the nucleotide at position 410 is a cytosine (C)(SEQ ID NO:74). In this embodiment, the amino acid at position 104 is aspartate (D)(SEQ ID NO:75).

The presence of a methionine residue at amino acid residue positions 6, 170, 192, and 210 of SEQ ID NO:2 indicates that there can be alternative forms of human TANGO 204 of 259 amino acids of SEQ ID NO:2 (SEQ ID NO:76), 95 amino acids of SEQ ID NO:2 (SEQ ID NO:77), 73 amino acids of SEQ ID NO:2 (SEQ ID NO:78), and 55 amino acids of SEQ ID NO:2 (SEQ ID NO:79), respectively.

Another embodiment of the invention includes isolated nucleic acid molecules comprising a polynucleotide having a nucleotide sequence encoding the polypeptide having the human TANGO 204 amino acid sequence in SEQ ID NO:2, but lacking the N-terminal methionine residue. In this embodiment, the nucleotide sequence of human TANGO 204, nucleotides 102-890 (SEQ ID NO:80), encodes the human TANGO 204 amino acid sequence from amino acids 2-264 (SEQ ID NO:81).

The signal peptide prediction program SIGNALP (Nielsen et al. (1997) *Protein Engineering* 10:1-6) predicted that human TANGO 204 includes a 20 amino acid signal peptide (amino acid 1 to about amino acid 20 of SEQ ID NO:2; SEQ ID NO:22) preceding the mature human TANGO 204 protein (corresponding to about amino acid 21 to amino acid 264 of SEQ ID NO:2; SEQ ID NO:23).

Human TANGO 204 that has not been post-translationally modified is predicted to have a molecular weight of 29.6 kDa prior to cleavage of its signal peptide and a molecular weight of 27.3 kDa subsequent to cleavage of its signal peptide.

Human TANGO 204 includes a somatomedin B domain at amino acids 18-75 of SEQ ID NO:2 (SEQ ID NO:36). Figure 3 depicts an alignment of the somatomedin B domain of human TANGO 204 with a consensus somatomedin B domain derived from a hidden Markov model (SEQ ID NO:37). Human TANGO 204 also includes a thrombospondin type I domain at amino acids 78-221 of SEQ ID NO:2 (SEQ ID NO:38). Figure 4 depicts an alignment of the thrombospondin type I domain of human TANGO 204 with a consensus thrombospondin type I domain derived from a hidden Markov model (SEQ ID NO:39).

An N-glycosylation site is present at amino acids 227-230 of SEQ ID NO:2. A cAMP and cGMP-dependent protein kinase phosphorylation site is present at amino acids 97-100 of SEQ ID NO:2. Protein kinase C phosphorylation sites are present at amino acids 93-95, 214-216, and 243-245 of SEQ ID NO:2. A casein kinase II phosphorylation site is present at amino acids 161-164 of SEQ ID NO:2. N-myristoylation sites are present at amino acids 17-22, 48-53, 129-134, and 236-241 of SEQ ID NO:2. A growth factor and cytokine receptor family signature sequence is present at amino acids 78-84 of SEQ ID NO:2. A somatomedin B domain signature sequence is present at amino acids 50-70 of SEQ ID NO:2.

Clone Athu204c, which encodes human TANGO 204, was deposited as fthv204c with the American Type Culture Collection (10801 University Boulevard, Manassas, VA 20110-2209) on April 2, 1999 and assigned Accession Number 207192. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. This deposit was made merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

Figure 2 depicts a hydropathy plot of human TANGO 204. Relatively hydrophobic regions are above the horizontal line, and relatively hydrophilic regions are below the horizontal line. The cysteine residues (cys) are indicated by short vertical lines just below the hydropathy trace. The hydropathy plot of Figure 2 indicates that human TANGO 204 has a signal sequence at its amino terminus and a hydrophobic region at its carboxy terminus, suggesting that TANGO 204 is a membrane-associated protein.

TANGO 204 is likely membrane-associated through its hydrophobic carboxy-terminus. The last nine amino acids of human TANGO 204 (amino acids 256-264 of SEQ

ID NO:2 (SEQ ID NO:178) are very hydrophobic. Further, there are two pairs of basic residues near the hydrophobic C-terminus (KK at amino acids 245-246 of SEQ ID NO:2 and RR at amino acids 248-249 of SEQ ID NO:2). These residues can serve as proteolytic cleavage sites. Thus, cleavage at either pair of basic residues can release a soluble form of TANGO 204 (amino acid 20-244, 20-245, 20-246, 20-287, 20-288, or 20-249 of SEQ ID NO:2). In addition, there is a RRR sequence at amino acids amino acids 97-99 of SEQ ID NO:2, and proteolytic cleavage at this sequence can release a soluble form of TANGO 204 (amino acids 20-96, 20-97, 20-98, or 20-99 of SEQ ID NO:2). The presence of a somatomedin B domain sequence within human TANGO 204 is consistent with TANGO 204 being a membrane-associated protein.

The human TANGO 204 gene maps to chromosome 8q between D8S257 and D8S508 based on the homology between a portion of human TANGO 204 and Genbank Accession Number G25656, which is reported to map to this position.

Mouse TANGO 204

A mouse homolog of human TANGO 204 was identified. A cDNA encoding murine TANGO 204 was identified by analyzing the sequences of clones present in a stimulated mouse osteoblast cDNA library.

This analysis led to the identification of a clone, Atmoa043g03, encoding full-length murine TANGO 204. The cDNA of this clone is 1294 nucleotides long (Figures 5A-5B; SEQ ID NO:13). The 792 nucleotide open reading frame of this cDNA, nucleotides 81-872 of SEQ ID NO:13 (SEQ ID NO:15), encodes a 264 amino acid protein (Figures 5A-5B; SEQ ID NO:14).

In one embodiment of a nucleotide sequence of mouse TANGO 204 the nucleotide at position 152 is a guanine (G)(SEQ ID NO:13). In this embodiment, the amino acid at position 24 is glutamate (E)(SEQ ID NO:14). In another embodiment of a nucleotide sequence of mouse TANGO 204, the nucleotide at position 152 is a cytosine (C)(SEQ ID NO:84). In this embodiment, the amino acid at position 24 is aspartate (D)(SEQ ID NO:85). In another embodiment of a nucleotide sequence of mouse TANGO 204, the nucleotide at position 392 is an adenine (A)(SEQ ID NO:13). In this embodiment, the amino acid at position 104 is a glutamate (E)(SEQ ID NO:14). In another embodiment of a nucleotide sequence of mouse TANGO 204, the nucleotide at

position 392 is a cytosine (C)(SEQ ID NO:86). In this embodiment, the amino acid at position 104 is aspartate (D)(SEQ ID NO:87). In another embodiment of a nucleotide sequence of mouse TANGO 204, the nucleotide at position 425 is an adenine (A)(SEQ ID NO:13). In this embodiment, the amino acid at position 116 is a glutamate (E)(SEQ ID NO:14). In another embodiment of a nucleotide sequence of mouse TANGO 204, the nucleotide at position 425 is a cytosine (C)(SEQ ID NO:88). In this embodiment, the amino acid at position 116 is aspartate (D)(SEQ ID NO:89).

The presence of a methionine residue at amino acid residue positions 6, 170, 192, and 210 of SEQ ID NO:14 indicates that there can be alternative forms of mouse TANGO 204 of 259 amino acids of SEQ ID NO:14 (SEQ ID NO:90), 95 amino acids of SEQ ID NO:14 (SEQ ID NO:91), 73 amino acids of SEQ ID NO:14 (SEQ ID NO:92), and 55 amino acids of SEQ ID NO:14 (SEQ ID NO:93), respectively.

Another embodiment of the invention includes isolated nucleic acid molecules comprising a polynucleotide having a nucleotide sequence encoding the polypeptide having the mouse TANGO 204 amino acid sequence in SEQ ID NO:14, but lacking the N-terminal methionine residue. In this embodiment, the nucleotide sequence of mouse TANGO 204, nucleotides 84-872 (SEQ ID NO:94), encodes the mouse TANGO 204 amino acid sequence comprising amino acids 2-264 (SEQ ID NO:95).

The signal peptide prediction program SIGNALP (Nielsen et al. (1997) *Protein Engineering* 10:1-6) predicted that murine TANGO 204 includes a 20 amino acid signal peptide (amino acid 1 to about amino acid 20 of SEQ ID NO:14; SEQ ID NO:30) preceding the mature murine TANGO 204 protein (corresponding to about amino acid 21 to amino acid 264 of SEQ ID NO:14; SEQ ID NO:31).

Murine TANGO 204 that has not been post-translationally modified is predicted to have a molecular weight of 29.5 kDa prior to cleavage of its signal peptide and a molecular weight of 27.2 kDa subsequent to cleavage of its signal peptide.

Murine TANGO 204 includes a somatomedin B domain at amino acids 18-75 of SEQ ID NO:14 (SEQ ID NO:40) and a thrombospondin type I domain at amino acids 78-121 of SEQ ID NO:14 (SEQ ID NO:41).

In situ tissue screening was performed on mouse adult and embryonic tissue to analyze the expression of mouse TANGO 204 mRNA. In summary, embryonic expression was observed in a number of tissues and organs. Most noticeable was the

expression in the eye, lung, stomach, intestine, and the tissue just under the skin in the feet which outlines the digits. Expression was also associated with some developing bone and cartilage structures such as the ear, nose, and spinal column. Expression decreased to background levels in most of these tissue and was observed in only a few adult tissues; eye, kidney, and adrenal gland.

Human and murine TANGO 204 sequences exhibit considerable similarity at the protein, nucleic acid, and open reading frame levels. An alignment (made using the ALIGN software (Myers and Miller (1989) CABIOS, ver. 2.0); BLOSUM 62 scoring matrix; gap penalties -12/-4), reveals a protein identity of 89.4%. The human and murine TANGO 204 full length cDNAs are 78.4 % identical, as assessed using the same software and parameters as indicated. In the respective ORFs, calculated in the same fashion as the full length cDNAs, human and murine TANGO 204 are 87.5% identical. The nucleotide sequence and amino acid sequence alignments of human and murine TANGO 204 can be found in Figures 6A-6B and Figure 7, respectively.

The mouse TANGO 204 gene was mapped to mouse using the Genebridge 4 Radiation hybrid mapping panel with gacaagctgcattcaaagcttcc (SEQ ID NO:82) as the forward primer and ctggagcacatggtagtgattc (SEQ ID NO:83) as the reverse primer. The mouse TANGO 204 gene maps to chromosome 1. Flanking markers for this region are D1Mit430 and D1Mit119. Mapping by synteny reveals that human TANGO 204 maps to human chromosome 8q. The CCAL1 (chondrocalcinosis 1) locus also maps to this region of the human chromosome. The OPRK (opiate receptor) gene also maps to this region of the human chromosome. The tb (tumbler), fz (fuzzy) loci also map to this region of the mouse chromosome. The tb (tumbler), fz (fuzzy) genes also map to this region of the mouse chromosome.

Clone Atmoa043g03, which encodes murine TANGO 204, was deposited as Atmoa43g3 with the American Type Culture Collection (10801 University Boulevard, Manassas, VA 20110-2209) on April 2, 1999 and assigned Accession Number 207189. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. This deposit was made merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

Use of TANGO 204 Nucleic Acids, Polypeptides, and Modulators Thereof

TANGO 204 polypeptides, nucleic acids, and modulators thereof, can be used to modulate the function, morphology, proliferation and/or differentiation of cells in the tissues in which it is expressed. TANGO 204 includes a thrombospondin type 1 domain. Known proteins having this domain play a role in blood coagulation, cellular proliferation, cellular adhesion, migration of tumor cells, migration of normal cells, and angiogenesis. The thrombospondin type 1 domain can mediate interaction with matrix macromolecules, including heparan sulfate, proteoglycans, fibronectin, laminin, and collagen. TANGO 204 polypeptides, nucleic acids, and modulators thereof can be used to treat disorders of blood clotting, angiogenesis (e.g., to reduce tumor growth by inhibiting angiogenesis or promote wound healing by stimulating angiogenesis), and cancer. TANGO 204 polypeptides, nucleic acids, and modulators thereof can also be used to treat connective tissue disorders (Marfan syndrome and osteogenesis imperfecta). TANGO 204 includes a somatomedin B domain. Known proteins having this domain are involved in regulation of plasminogen activator inhibitor, a protein which regulates activity of plasmin, a protein involved in ovulation, angiogenesis, neoplasia, wound healing, embryonic development, and inflammation. Thus, TANGO 204 polypeptides, nucleic acids, and modulators thereof can also be used to treat disorders of ovulation. In addition, such molecules can be used to treat disorders associated with proteases in cardiovascular tissue, disorders of complement activation, and disorders of fibrinolysis.

With respect to angiogenisis in particular, angiogenesis is also involved in pathological conditions including the growth and metastasis of tumors. In fact, tumor growth and metastasis have been shown to be dependent on the formation of new blood vessels. Accordingly, TANGO 204 polypeptides, nucleic acids and/or modulators thereof can be used to modulate angiogenesis in proliferative disorders such as cancer, (*e.g.*, fibrosarcoma, myxosarcoma, liposarcoma, chondrosarcoma, osteogenic sarcoma, chordoma, angiosarcoma, endotheliosarcoma, lymphangiosarcoma, lymphangioendotheliosarcoma, synovioma, mesothelioma, Ewing's tumor, leimyosarcoma, rhabdotheliosarcoma, colon sarcoma, pancreatic cancer, breast cancer, ovarian cancer, prostate cancer, squamous cell carcinoma, basal cell carcinoma, adenocarcinoma, sweat gland carcinoma, sebaceous gland carcinoma, papillary carcinoma, papillary adenocarcinomas, cystadenocarcinoma, medullary carcinoma,

bronchogenic carcinoma, renal cell carcinoma, hematoma, bile duct carcinoma, melanoma, choriocarcinoma, semicoma, embryonal carcinoma, Wilms' tumor, cervical cancer, testicular tumor, lung carcinoma, small cell carcinoma, bladder carcinoma, epithelial carcinoma, glioma, astrocytoma, medulloblastoma, craniopharyngioma, ependynoma, pinealoma, hemangioblastoma, and retinoblastoma.

TANGO 204 polypeptides, nucleic acids, and modulators thereof, can be used to modulate the function, morphology, proliferation and/or differentiation of cells in the tissues in which it is expressed. Such molecules can be used to treat disorders associated with abnormal or aberrant metabolism or function of cells in the tissues in which it is expressed. Tissues in which TANGO 204 is expressed include, for example, eye, stomach, intestine, cortex adrenal gland, kidney, developing bone and cartilage structures such as the ear, nose, and spinal column, and the pericardium surrounding the heart.

In another example, because TANGO 204 is expressed in the pericardium surrounding the heart TANGO 201 polypeptides, nucleic acids, or modulators thereof, can be used to treat cardiovascular disorders, such as ischemic heart disease (e.g., angina pectoris, myocardial infarction, and chronic ischemic heart disease), hypertensive heart disease, pulmonary heart disease, valvular heart disease (e.g., rheumatic fever and rheumatic heart disease, endocarditis, mitral valve prolapse, and aortic valve stenosis), congenital heart disease (e.g., valvular and vascular obstructive lesions, atrial or ventricular septal defect, and patent ductus arteriosus), or myocardial disease (e.g., myocarditis, congestive cardiomyopathy, and hypertrophic cariomyopathy).

Because TANGO 204 is expressed in the kidney, the TANGO 204 polypeptides, nucleic acids and/or modulators thereof can be used to modulate the function, morphology, proliferation and/or differentiation of cells in the tissues in which it is expressed. Such molecules can also be used to treat disorders associated with abnormal or aberrant metabolism or function of cells in the tissues in which it is expressed. Such molecules can be used to treat or modulate renal (kidney) disorders, such as glomerular diseases (e.g., acute and chronic glomerulonephritis, rapidly progressive glomerulonephritis, nephrotic syndrome, focal proliferative glomerulonephritis, glomerular lesions associated with systemic disease, such as systemic lupus erythematosus, Goodpasture's syndrome, multiple myeloma, diabetes, neoplasia, sickle cell disease, and chronic inflammatory diseases), tubular diseases (e.g., acute tubular

necrosis and acute renal failure, polycystic renal diseasemedullary sponge kidney, medullary cystic disease, nephrogenic diabetes, and renal tubular acidosis), tubulointerstitial diseases (e.g., pyelonephritis, drug and toxin induced tubulointerstitial nephritis, hypercalcemic nephropathy, and hypokalemic nephropathy) acute and rapidly progressive renal failure, chronic renal failure, nephrolithiasis, vascular diseases (e.g., hypertension and nephrosclerosis, microangiopathic hemolytic anemia, atheroembolic renal disease, diffuse cortical necrosis, and renal infarcts), or tumors (e.g., renal cell carcinoma and nephroblastoma).

As TANGO 204 exhibits expression in the small intestine, TANGO 204 polypeptides, nucleic acids, or modulators thereof, can be used to treat intestinal disorders, such as ischemic bowel disease, infective enterocolitis, Crohn's disease, benign tumors, malignant tumors (e.g., argentaffinomas, lymphomas, adenocarcinomas, and sarcomas), malabsorption syndromes (e.g., celiac disease, tropical sprue, Whipple's disease, and abetalipoproteinemia), obstructive lesions, hernias, intestinal adhesions, intussusception, or volvulus.

As murine TANGO 204 was originally identified in an osteoblast cDNA library, TANGO 204 nucleic acids, proteins, and modulators thereof can be used to modulate the proliferation, activation, development, differentiation, and/or function of osteoblasts. Thus, TANGO 204 nucleic acids, proteins, and modulators thereof can be used to modulate the proliferation, differentiation, and/or function of bone and cartilage cells, *e.g.*, chondrocytes and osteoblasts, and to treat bone and/or cartilage associated diseases or disorders. Examples of bone and/or cartilage diseases and disorders include bone and/or cartilage injury due to for example, trauma (*e.g.*, bone breakage, cartilage tearing), degeneration (*e.g.*, osteoporosis), degeneration of joints, *e.g.*, arthritis, *e.g.*, osteoarthritis, and bone wearing.

As human TANGO 204 was originally identified in a lung cDNA library, human TANGO 204 nucleic acids, proteins, and modulators thereof can be used to modulate the proliferation, activation, development, differentiation, and/or function of lung cells. Thus, TANGO 204 polypeptides, nucleic acids, or modulators thereof, can be used to treat pulmonary (lung) disorders, such as atelectasis, pulmonary congestion or edema, chronic obstructive airway disease (e.g., emphysema, chronic bronchitis, bronchial asthma, and bronchiectasis), diffuse interstitial diseases (e.g., sarcoidosis,

pneumoconiosis, hypersensitivity pneumonitis, Goodpasture's syndrome, idiopathic pulmonary hemosiderosis, pulmonary alveolar proteinosis, desquamative interstitial pneumonitis, chronic interstitial pneumonia, fibrosing alveolitis, hamman-rich syndrome, pulmonary eosinophilia, diffuse interstitial fibrosis, Wegener's granulomatosis, lymphomatoid granulomatosis, and lipid pneumonia), or tumors (e.g., bronchogenic carcinoma, bronchiolovlveolar carcinoma, bronchial carcinoid, hamartoma, and mesenchymal tumors).

In another example, TANGO 204 polypeptides, nucleic acids, or modulators thereof, can be used to treat disorders of the adrenal cortex, such as hypoadrenalism (e.g., primary chronic or acute adrenocortical insufficiency, and secondary adrenocortical insufficiency), hyperadrenalism (Cushing's syndrome, primary hyperaldosteronism, adrenal virilism, and adrenal hyperplasia), or neoplasia (e.g., adrenal adenoma and cortical carcinoma).

Human TANGO 206

A cDNA encoding human TANGO 206 was identified by analyzing the sequences of clones present in a human osteoblast cDNA library.

This analysis led to the identification of a clone, Athoc49b12, encoding full-length human TANGO 206. The cDNA of this clone is 1840 nucleotides long (Figures 8A-8B; SEQ ID NO:4). The 1260 nucleotide open reading frame of this cDNA, nucleotides 99-1358 of SEQ ID NO:4 (SEQ ID NO:6), encodes a 420 amino acid protein (Figures 8A-8B; SEQ ID NO:5).

In one embodiment of a nucleotide sequence of human TANGO 206 the nucleotide at position 281 is a guanine (G)(SEQ ID NO:4). In this embodiment, the amino acid at position 61 is glutamate (E)(SEQ ID NO:5). In another embodiment of a nucleotide sequence of human TANGO 206, the nucleotide at position 281 is a cytosine (C)(SEQ ID NO:96). In this embodiment, the amino acid at position 61 is aspartate (D)(SEQ ID NO:97) In another embodiment of a nucleotide sequence of human TANGO 206, the nucleotide at position 326 is a guanine (G)(SEQ ID NO:4). In this embodiment, the amino acid at position 76 is a glutamate (E)(SEQ ID NO:5). In another embodiment of a nucleotide sequence of human TANGO 206, the nucleotide at position 326 is a cytosine (C)(SEQ ID NO:98). In this embodiment, the amino acid at position 76 is

aspartate (D)(SEQ ID NO:99). In another embodiment of a nucleotide sequence of human TANGO 206, the nucleotide at position 329 is an adenine (A)(SEQ ID NO:4). In this embodiment, the amino acid at position 77 is a glutamate (E)(SEQ ID NO:5). In another embodiment of a nucleotide sequence of human TANGO 206, the nucleotide at position 329 is a cytosine (C)(SEQ ID NO:100). In this embodiment, the amino acid at position 77 is aspartate (D)(SEQ ID NO:101).

The presence of a methionine residue at amino acid residue positions 282, 339, 358, 369, and 400 of SEQ ID NO:5 indicates that there can be alternative forms of human TANGO 206 of 139 amino acids of SEQ ID NO:5 (SEQ ID NO:102), 82 amino acids of SEQ ID NO:5 (SEQ ID NO:103), 63 amino acids of SEQ ID NO:5 (SEQ ID NO:104), 52 amino acids of SEQ ID NO:5 (SEQ ID NO:105), and 21 amino acids of SEQ ID NO:5 (SEQ ID NO:106), respectively.

Another embodiment of the invention includes isolated nucleic acid molecules comprising a polynucleotide having a nucleotide sequence encoding the polypeptide having the human TANGO 206 amino acid sequence in SEQ ID NO:5, but lacking the N-terminal methionine residue. In this embodiment, the nucleotide sequence of human TANGO 206, nucleotides 102-1358 (SEQ ID NO:107), encodes the human TANGO 206 amino acid sequence comprising amino acids 2-420 (SEQ ID NO:108).

The signal peptide prediction program SIGNALP (Nielsen et al. (1997) *Protein Engineering* 10:1-6) predicted that human TANGO 206 includes a 29 amino acid signal peptide (amino acid 1 to about amino acid 29 of SEQ ID NO:5; SEQ ID NO:24) preceding the mature human TANGO 206 protein (corresponding to about amino acid 30 to amino acid 420 of SEQ ID NO:5; SEQ ID NO:25).

Human TANGO 206 is a transmembrane protein having a first extracellular domain which extends from about amino acid 30 to about amino acid 362 of SEQ ID NO:5 (SEQ ID NO:42), a first transmembrane domain which extends from about amino acid 363 to about amino acid 379 of SEQ ID NO:5 (SEQ ID NO:43), a cytoplasmic domain which extends from about amino acid 380 to about amino acid 386 of SEQ ID NO:5 (SEQ ID NO:44), a second transmembrane domain which extends from about amino acid 387 to about amino acid 405 of SEQ ID NO:5 (SEQ ID NO:45), and a second extracellular domain which extends from about amino acid 406 to amino acid 420 of SEQ ID NO:5 (SEQ ID NO:46).

Alternatively, in another embodiment, a human TANGO 206 is a transmembrane protein having a first cytoplasmic domain which extends from about amino acid 30 to about amino acid 362 of SEQ ID NO:5 (SEQ ID NO:179), a first transmembrane domain which extends from about amino acid 363 to about amino acid 379 of SEQ ID NO:5 (SEQ ID NO:43), an extracellular domain which extends from about amino acid 380 to about amino acid 386 of SEQ ID NO:5 (SEQ ID NO:180), a second transmembrane domain which extends from about amino acid 387 to about amino acid 405 of SEQ ID NO:5 (SEQ ID NO:45), and a second cytoplasmic domain which extends from about amino acid 406 to amino acid 420 of SEQ ID NO:5 (SEQ ID NO:181).

Human TANGO 206 includes a laminin EGF-like domain at amino acids 168-211 of SEQ ID NO:5 (SEQ ID NO:47). Figure 10 depicts an alignment of the laminin EGF-like domain of human TANGO 206 with a laminin EGF-like domain derived from a hidden Markov model (SEQ ID NO:48).

Human TANGO 206 that has not been post-translationally modified is predicted to have a molecular weight of 45.4 kDa prior to cleavage of its signal peptide and a molecular weight of 42.1 kDa subsequent to cleavage of its signal peptide.

N-glycosylation sites are present at amino acids 79-82 and 205-208 of SEQ ID NO:5. A cAMP and cGMP-dependent protein kinase phosphorylation site is present at amino acids 290-293 of SEQ ID NO:5. Protein kinase C phosphorylation sites are present at amino acids 48-50, 63-65, 138-140, 159-161, 406-408, and 409-411 of SEQ ID NO:5. Casein kinase II phosphorylation sites are present at amino acids 63-66, 73-76, 99-102, 222-225, and 359-362 of SEQ ID NO:5. N-myristoylation sites are present at amino acids 8-13, 51-56, 59-64, 69-74, 167-172, 173-178, 188-193, 250-255, 267-272, 280-285, 326-331, 372-377, and 395-400 of SEQ ID NO:5. An aspartic acid and asparagine hydroxylation site is present at amino acids 321-332 of SEQ ID NO:5. An EGF-like domain cysteine pattern signature is present at amino acids 181-192 of SEQ ID NO:5.

Clone Athoc49b12, which encodes human TANGO 206, was deposited as EpT206 with the American Type Culture Collection (10801 University Boulevard, Manassas, VA 20110-2209) on April 21, 1999 and assigned Accession Number 207223. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent

Procedure. This deposit was made merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

Figure 9 depicts a hydropathy plot of human TANGO 206. Relatively hydrophobic regions are above the horizontal line, and relatively hydrophilic regions are below the horizontal line. The cysteine residues (cys) are indicated by short vertical lines just below the hydropathy trace. The hydropathy plot of Figure 9 indicates the presence of a signal sequence at the amino-terminus of human TANGO 206 and two transmembrane domains within human TANGO 206, suggesting that human TANGO 206 is a transmembrane protein.

Northern analysis of human TANGO 206 mRNA expression revealed strong expression in the heart, moderate expression in the skeletal muscle and weak expression in the kidney, brain, and placenta.

The human TANGO 206 gene maps to chromosome 3 between D3S3591 and D3S1283 based on the homology between a portion of human TANGO 206 and Genbank Accession Number G06979 (human STS WI-8719), which is reported to map to this position.

Mouse TANGO 206

A cDNA encoding murine TANGO 206 was identified by analyzing the sequences of clones present in a mouse bone marrow cDNA library.

This analysis led to the identification of a clone, AtmMa206, encoding full-length murine TANGO 206. The cDNA of this clone is 2093 nucleotides long (Figures 11A-11B; SEQ ID NO:16). The 1260 nucleotide open reading frame of this cDNA, nucleotides 332-1591 of SEQ ID NO:16 (SEQ ID NO:18), encodes a 420 amino acid protein (Figures 11A-11B; SEQ ID NO:17).

In one embodiment of a nucleotide sequence of mouse TANGO 206, the nucleotide at position 457 is a guanine (G)(SEQ ID NO:16). In this embodiment, the amino acid at position 42 is glutamate (E)(SEQ ID NO:17). In another embodiment of a nucleotide sequence of mouse TANGO 206, the nucleotide at position 457 is a cytosine (C)(SEQ ID NO:109). In this embodiment, the amino acid at position 42 is aspartate (D)(SEQ ID NO:110) In another embodiment of a nucleotide sequence of mouse TANGO 206, the nucleotide at position 514 is a guanine (G)(SEQ ID NO:16). In this

embodiment, the amino acid at position 61 is a glutamate (E)(SEQ ID NO:17). In another embodiment of a nucleotide sequence of mouse TANGO 206, the nucleotide at position 514 is a cytosine (C)(SEQ ID NO:111). In this embodiment, the amino acid at position 61 is aspartate (D)(SEQ ID NO:112). In another embodiment of a nucleotide sequence of mouse TANGO 206, the nucleotide at position 559 is an adenine (A)(SEQ ID NO:16). In this embodiment, the amino acid at position 76 is a glutamate (E)(SEQ ID NO:17). In another embodiment of a nucleotide sequence of mouse TANGO 206, the nucleotide at position 559 is a cytosine (C)(SEQ ID NO:113). In this embodiment, the amino acid at position 76 is aspartate (D)(SEQ ID NO:114).

The presence of a methionine residue at positions 282, 358, 363, 369, and 400 of SEQ ID NO:17 indicates that there can be alternative forms of mouse TANGO 206 of 139 amino acids of SEQ ID NO:17 (SEQ ID NO:115), 63 amino acids of SEQ ID NO:17 (SEQ ID NO:116), 58 amino acids of SEQ ID NO:17 (SEQ ID NO:117), 52 amino acids of SEQ ID NO:17 (SEQ ID NO:118), and 21 amino acids of SEQ ID NO:17 (SEQ ID NO:119), respectively.

Another embodiment of the invention includes isolated nucleic acid molecules comprising a polynucleotide having a nucleotide sequence encoding the polypeptide having the mouse TANGO 206 amino acid sequence in SEQ ID NO:17, but lacking the N-terminal methionine residue. In this embodiment, the nucleotide sequence of mouse TANGO 206, nucleotides 335-1591 (SEQ ID NO:120), encodes the mouse TANGO 206 amino acid sequence from amino acids 2-420 (SEQ ID NO:121).

The signal peptide prediction program SIGNALP (Nielsen et al. (1997) *Protein Engineering* 10:1-6) predicted that murine TANGO 206 includes a 29 amino acid signal peptide (amino acid 1 to about amino acid 29 of SEQ ID NO:17; SEQ ID NO:32) preceding the mature murine TANGO 206 protein (corresponding to about amino acid 30 to amino acid 420 of SEQ ID NO:17; SEQ ID NO:33).

Murine TANGO 206 is a transmembrane protein having a first extracellular domain which extends from about amino acid 30 to about amino acid 362 of SEQ ID NO:17 (SEQ ID NO:65), a first transmembrane domain which extends from about amino acid 363 to about amino acid 379 of SEQ ID NO:17 (SEQ ID NO:66), a cytoplasmic domain which extends from about amino acid 380 to about amino acid 386 of SEQ ID NO:17 (SEQ ID NO:67), a second transmembrane domain which extends from about

amino acid 387 to about amino acid 405 of SEQ ID NO:17 (SEQ ID NO:69), and a second extracellular domain which extends from about amino acid 406 to about amino acid 420 of SEQ ID NO:17 (SEQ ID NO:68).

Alterantively, murine TANGO 206 is a transmembrane protein having a first cytoplasmic domain which extends from about amino acid 30 to about amino acid 362 of SEQ ID NO:17 (SEQ ID NO:183), a first transmembrane domain which extends from about amino acid 363 to about amino acid 379 of SEQ ID NO:17 (SEQ ID NO:66), an extracellular domain which extends from about amino acid 380 to about amino acid 386 of SEQ ID NO:17 (SEQ ID NO:184), a second transmembrane domain which extends from about amino acid 387 to about amino acid 405 of SEQ ID NO:17 (SEQ ID NO:69), and a second cytoplasmic domain which extends from about amino acid 406 to about amino acid 420 of SEQ ID NO:17 (SEQ ID NO:185).

Murine TANGO 206 that has not been post-translationally modified is predicted to have a molecular weight of 45.7 kDa prior to cleavage of its signal peptide and a molecular weight of 42.4 kDa subsequent to cleavage of its signal peptide.

Mouse TANGO 206 includes a laminin EGF-like domain at amino acids 168-211 of SEQ ID NO:17 (SEQ ID NO:49) and two EGF-like domains, one at amino acids 155-192 of SEQ ID NO:17 (SEQ ID NO:182) and one at amino acids 309-343 of SEQ ID NO:17 (SEQ ID NO:50).

In situ tissue screening was performed on mouse adult and embryonic tissue to analyze the expression of mouse TANGO 206 mRNA. In summary, expression during embryogenesis was observed ubiquitously in the central nervous system of the ages examined. It was also observed in the eye and the large ganglion of the head. Expression was also observed in the liver from E13.5 to E15.5. Expression pattern was multifocal in a pattern suggestive of megakaryocytes or haemopoietic islands. Expression was also observed in the skin of the earlier embryonic ages. Adult expression was observed ubiquitously in the brain and grey matter of the spinal cord. The adrenal gland and small intestine also had moderate to strong expression.

Human and murine TANGO 206 sequences exhibit considerable similarity at the protein, nucleic acid, and open reading frame levels. An alignment (made using the ALIGN software (Myers and Miller (1989) CABIOS, ver. 2.0); BLOSUM 62 scoring matrix; gap penalties -12/-4), reveals a protein identity of 91.4%. The human and murine

TANGO 206 full length cDNAs are 84% identical, as assessed using the same software and parameters as indicated. In the respective ORFs, calculated in the same fashion as the full length cDNAs, human and murine TANGO 206 are 89% identical. The nucleotide sequence and amino acid sequence alignments of human and murine TANGO 206 can be found in Figures 12A-12C and Figure 13, respectively.

Clone AtmMa206, which encodes murine TANGO 206, was deposited as EpTm206 with the American Type Culture Collection (10801 University Boulevard, Manassas, VA 20110-2209) on April 21, 1999 and assigned Accession Number 207221. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. This deposit was made merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

Use of TANGO 206 Nucleic Acids, Polypeptides, and Modulators Thereof

TANGO 206 polypeptides, nucleic acids, and modulators thereof, can be used to modulate the function, morphology, proliferation and/or differentiation of cells in the tissues in which it is expressed. TANGO 206 includes an laminin EGF domain and an EGF-like domain. Proteins having such domains play a role in a wide variety of biological processes, including cholesterol uptake, blood coagulation, specification of cell fate. TANGO 206 polypeptides, nucleic acids, and modulators thereof can be used to modulate cell proliferation, morphogenesis, tissue repair and renewal, terminal differentiation, cell survival, and cell migration. They can be used to treat cancer, promote would healing (e.g., of the skin, cornea, or digestive mucosa), treat familia hypercholesterolemia, treat hemophilia B, treat Marfan syndrome, and treat protein S deficiency, and modulate an allergic or inflammatory response. TANGO 206 polypeptides, nucleic acids, and modulators thereof can be used to modulate acid secretion, modulate tropic effects on gastrointestinal mucosa, modulate mucosal adaptation, and modulate gastroduodenal cell migration and proliferation. Thus, such molecules can be used to protect gastric mucosa against injury and promote gastroduodenal ulcer healing.

As human TANGO 206 was originally found in a LPS stimulated human primary osteoblast library, TANGO 206 nucleic acids, proteins, and modulators thereof

can be used to modulate the proliferation, differentiation, and/or function of cells that form bone matrix, e.g., osteoblasts and osteoclasts, and can be used to modulate the formation of bone matrix. Thus A259 nucleic acids, proteins, and modulators thereof can be used to treat cartilage and bone associated diseases and disorders, and can play a role in bone growth, formation, and remodeling. Examples of cartilage and bone associated diseases and disorders include e.g., bone cancer, achondroplasia, myeloma, fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma, and osteoporosis.

As murine TANGO 206 was originally found in a bone marrow library, TANGO 206 nucleic acids, proteins, and modulators thereof can be used to modulate the proliferation, differentiation, and/or function of cells that appear in the bone marrow, e.g., stem cells (e.g., hematopoietic stem cells), and blood cells, e.g., erythrocytes, platelets, and leukocytes. Thus A259 nucleic acids, proteins, and modulators thereof can be used to treat bone marrow, blood, and hematopoietic associated diseases and disorders, e.g., acute myeloid leukemia, hemophilia, leukemia, anemia (e.g., sickle cell anemia), and thalassemia.

TANGO 206 polypeptides, nucleic acids, and modulators thereof, can be used to modulate the function, morphology, proliferation and/or differentiation of cells in the tissues in which it is expressed. Such molecules can be used to treat disorders associated with abnormal or aberrant metabolism or function of cells in the tissues in which it is expressed. Tissues in which TANGO 206 is expressed include, for example, heart, brain, skeletal muscle, placenta, CNS, liver, small intestine, adrenal gland, and the kidney.

In another example, TANGO 206 polypeptides, nucleic acids, or modulators thereof, can be used to treat disorders of the brain, such as cerebral edema, hydrocephalus, brain herniations, iatrogenic disease (due to, e.g., infection, toxins, or drugs), inflammations (e.g., bacterial and viral meningitis, encephalitis, and cerebral toxoplasmosis), cerebrovascular diseases (e.g., hypoxia, ischemia, and infarction, intracranial hemorrhage and vascular malformations, and hypertensive encephalopathy), and tumors (e.g., neuroglial tumors, neuronal tumors, tumors of pineal cells, meningeal tumors, primary and secondary lymphomas, intracranial tumors, and medulloblastoma), and to treat injury or trauma to the brain.

In another example, TANGO 206 polypeptides, nucleic acids, or modulators thereof, can be used to treat pancreatic disorders, such as pancreatitis (e.g., acute

hemorrhagic pancreatitis and chronic pancreatitis), pancreatic cysts (e.g., congenital cysts, pseudocysts, and benign or malignant neoplastic cysts), pancreatic tumors (e.g., pancreatic carcinoma and adenoma), diabetes mellitus (e.g., insulin- and non-insulin-dependent types, impaired glucose tolerance, and gestational diabetes), or islet cell tumors (e.g., insulinomas, adenomas, Zollinger-Ellison syndrome, glucagonomas, and somatostatinoma).

As TANGO 206 exhibits expression in the heart, TANGO 206 nucleic acids, proteins, and modulators thereof can be used to treat cardiovascular disorders, such as ischemic heart disease (e.g., angina pectoris, myocardial infarction, and chronic ischemic heart disease), hypertensive heart disease, pulmonary heart disease, valvular heart disease (e.g., rheumatic fever and rheumatic heart disease, endocarditis, mitral valve prolapse, and aortic valve stenosis), congenital heart disease (e.g., valvular and vascular obstructive lesions, atrial or ventricular septal defect, and patent ductus arteriosus), or myocardial disease (e.g., myocarditis, congestive cardiomyopathy, and hypertrophic cardiomyopathy), atherosclerosis, hypertension, and angina pectoris.

In another example, TANGO 206 polypeptides, nucleic acids, or modulators thereof, can be used to treat hepatic (liver) disorders, such as jaundice, hepatic failure, hereditary hyperbiliruinemias (e.g., Gilbert's syndrome, Crigler-Naijar syndromes and Dubin-Johnson and Rotor's syndromes), hepatic circulatory disorders (e.g., hepatic vein thrombosis and portal vein obstruction and thrombosis) hepatitis (e.g., chronic active hepatitis, acute viral hepatitis, and toxic and drug-induced hepatitis) cirrhosis (e.g., alcoholic cirrhosis, biliary cirrhosis, and hemochromatosis), or malignant tumors (e.g., primary carcinoma, hepatoblastoma, and angiosarcoma).

In another example, TANGO 206 polypeptides, nucleic acids, or modulators thereof, can be used to treat renal (kidney) disorders, such as glomerular diseases (e.g., acute and chronic glomerulonephritis, rapidly progressive glomerulonephritis, nephrotic syndrome, focal proliferative glomerulonephritis, glomerular lesions associated with systemic disease, such as systemic lupus erythematosus, Goodpasture's syndrome, multiple myeloma, diabetes, neoplasia, sickle cell disease, and chronic inflammatory diseases), tubular diseases (e.g., acute tubular necrosis and acute renal failure, polycystic renal diseasemedullary sponge kidney, medullary cystic disease, nephrogenic diabetes, and renal tubular acidosis), tubulointerstitial diseases (e.g., pyelonephritis, drug and toxin

induced tubulointerstitial nephritis, hypercalcemic nephropathy, and hypokalemic nephropathy) acute and rapidly progressive renal failure, chronic renal failure, nephrolithiasis, vascular diseases (e.g., hypertension and nephrosclerosis, microangiopathic hemolytic anemia, atheroembolic renal disease, diffuse cortical necrosis, and renal infarcts), or tumors (e.g., renal cell carcinoma and nephroblastoma).

In another example, TANGO 206 polypeptides, nucleic acids, or modulators thereof, can be used to treat intestinal disorders, such as ischemic bowel disease, infective enterocolitis, Crohn's disease, benign tumors, malignant tumors (e.g., argentaffinomas, lymphomas, adenocarcinomas, and sarcomas), malabsorption syndromes (e.g., celiac disease, tropical sprue, Whipple's disease, and abetalipoproteinemia), obstructive lesions, hernias, intestinal adhesions, intussusception, or volvulus.

In another example, TANGO 206 polypeptides, nucleic acids, or modulators thereof, can be used to treat disorders of the adrenal cortex, such as hypoadrenalism (e.g., primary chronic or acute adrenocortical insufficiency, and secondary adrenocortical insufficiency), hyperadrenalism (Cushing's syndrome, primary hyperaldosteronism, adrenal virilism, and adrenal hyperplasia), or neoplasia (e.g., adrenal adenoma and cortical carcinoma).

Human TANGO 209

A cDNA encoding human TANGO 209 was identified by analyzing the sequences of clones present in a human osteoblast cDNA library.

This analysis led to the identification of a clone, Athoc22d3, encoding full-length human TANGO 209. The cDNA of this clone is 3117 nucleotides long (Figures 14A-14C; SEQ ID NO:7). The 1338 nucleotide open reading frame of this cDNA, nucleotides 194-1531 of SEQ ID NO:7 (SEQ ID NO:9), encodes a 446 amino acid protein (Figures 14A-14C; SEQ ID NO:8).

In one embodiment of a nucleotide sequence of human TANGO 209, the nucleotide at position 388 is an adenine (A)(SEQ ID NO:7). In this embodiment, the amino acid at position 65 is glutamate (E)(SEQ ID NO:8). In another embodiment of a nucleotide sequence of human TANGO 209, the nucleotide at position 388 is a cytosine (C)(SEQ ID NO:122). In this embodiment, the amino acid at position 65 is aspartate (D)(SEQ ID NO:123) In another embodiment of a nucleotide sequence of human

TANGO 209, the nucleotide at position 424 is a guanine (G)(SEQ ID NO:7). In this embodiment, the amino acid at position 77 is a glutamate (E)(SEQ ID NO:8). In another embodiment of a nucleotide sequence of human TANGO 209, the nucleotide at position 424 is a cytosine (C)(SEQ ID NO:124). In this embodiment, the amino acid at position 77 is aspartate (D)(SEQ ID NO:125). In another embodiment of a nucleotide sequence of human TANGO 209, the nucleotide at position 472 is an adenine (A)(SEQ ID NO:7). In this embodiment, the amino acid at position 93 is a glutamate (E)(SEQ ID NO:8). In another embodiment of a nucleotide sequence of human TANGO 209, the nucleotide at position 472 is a cytosine (C)(SEQ ID NO:126). In this embodiment, the amino acid at position 93 is aspartate (D)(SEQ ID NO:127).

The presence of a methionine residue at positions 324, and 410 of SEQ ID NO:8 indicates that there can be alternative forms of human TANGO 209 of 123 amino acids of SEQ ID NO:8 (SEQ ID NO:128), and 37 amino acids of SEQ ID NO:8 (SEQ ID NO:129), respectively.

Another embodiment of the invention includes isolated nucleic acid molecules comprising a polynucleotide having a nucleotide sequence encoding the polypeptide having the human TANGO 209 amino acid sequence in SEQ ID NO:8, but lacking the N-terminal methionine residue. In this embodiment, the nucleotide sequence of human TANGO 209, nucleotides 197-1531 (SEQ ID NO:130), encodes the human TANGO 209 amino acid sequence from amino acids 2-446 (SEQ ID NO:131).

The signal peptide prediction program SIGNALP (Nielsen et al. (1997) *Protein Engineering* 10:1-6) predicted that human TANGO 209 includes a 21 amino acid signal peptide (amino acid 1 to about amino acid 21 of SEQ ID NO:8; SEQ ID NO:26) preceding the mature human TANGO 209 protein (corresponding to about amino acid 22 to amino acid 446 of SEQ ID NO:8; SEQ ID NO:27).

Human TANGO 209 that has not been post-translationally modified is predicted to have a molecular weight of 49.7 kDa prior to cleavage of its signal peptide and a molecular weight of 47.3 kDa subsequent to cleavage of its signal peptide.

Human TANGO 209 includes thyroglobulin type 1 repeat domains at amino acids 109-153 of SEQ ID NO:8 (SEQ ID NO:51) and amino acids 237-281 of SEQ ID NO:8 (SEQ ID NO:52) . Figure 16 depicts an alignment of the thyroglobulin type 1 repeat domains of human TANGO 209 with a consensus thyroglobulin type 1 repeat

domain derived from a hidden Markov model (SEQ ID NO:53). Human TANGO 209 includes a Kazal-type serine protease inhibitor domain at amino acids 40-84 of SEQ ID NO:8 (SEQ ID NO:54). Figure 17 depicts an alignment of the Kazal-type serine protease inhibitor domain of human TANGO 209 with a consensus Kazal-type serine protease domain derived from a hidden Markov model (SEQ ID NO:55).

N-glycosylation sites are present at amino acids 206-209 and 362-365 of SEQ ID NO:8. In human TANGO 209, cAMP and cGMP-dependent protein kinase phosphorylation sites are present at amino acids 94-97, 380-383, 426-429 of SEQ ID NO:8. Protein kinase C phosphorylation sites are present at amino acids 150-152, 167-169, 208-210, 265-267, 273-275, 284-286, 335-337, 424-426, 429-431, and 438-440 of SEQ ID NO:8. Casein kinase II phosphorylation sites are present at amino acids 62-65, 156-159, 214-217, 222-225, 274-277, 315-318, 339-342, 346-349, 363-366, and 405-408 of SEQ ID NO:8. A tyrosine kinase phosphorylation site is present at amino acids 89-96 of SEQ ID NO:8. N-myristoylation sites are present at amino acids 143-148, 166-171, and 303-308 of SEQ ID NO:8. An amidation site is present at amino acids 367-370 of SEQ ID NO:8. EF-hand calcium-binding domains are present at amino acids 360-372 and 397-409 of SEQ ID NO:8. A thyroglobulin type-1 repeat signature is present at amino acids 109-138 of SEQ ID NO:8.

Clone Athoc22d3, which encodes human TANGO 209, was deposited as EpT209 with the American Type Culture Collection (10801 University Boulevard, Manassas, VA 20110-2209) on April 21, 1999 and assigned Accession Number 207223. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. This deposit was made merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

Figure 15 depicts a hydropathy plot of human TANGO 209. Relatively hydrophobic regions are above the horizontal line, and relatively hydrophilic regions are below the horizontal line. The cysteine residues (cys) are indicated by short vertical lines just below the hydropathy trace. The hydropathy plot of Figure 15 indicates that human TANGO 209 has a signal sequence at its amino terminus, suggesting that human TANGO 209 is a secreted protein.

Northern analysis of human TANGO 209 mRNA expression revealed very high expression in the heart, high expression in the skeletal muscle and pancreas, and moderate expression in the placenta, lung and kidney.

The human gene for TANGO 209 was mapped on radiation hybrid panels to the long arm of chromosome 6, in the region q26-27. Flanking markers for this region are ATA22G07 and WI-9405. The MLLT4 (myeloid/lymphoid or mixed lineage leukemia) locus also maps to this region of the human chromosome. The PLG (plasminogen), VIP (vasoactive intestinal peptide), LPA (apolipoprotein Lp), MLLT4 (myeloid/lymphoid or mixed lineage leukemia), and THBS2 (thrombospondin 2) genes also map to this region of the human chromosome. This region is syntenic to mouse chromosome 17. The qk (quaking), T (brachyury), and het (head tilt) loci also map to this region of the mouse chromosome. The plg (plasminogen), qk (quaking), and het (head tilt) genes also map to this region of the mouse chromosome.

Mouse TANGO 209

A cDNA encoding murine TANGO 209 was identified by analyzing the sequences of clones present in a mouse osteoblast cDNA library.

This analysis led to the identification of a clone, Atmoa99h11, encoding full-length murine TANGO 209. The cDNA of this clone is 2810 nucleotides long (Figures 18A-18C; SEQ ID NO:19). The 1341 nucleotide open reading frame of this cDNA, nucleotides 187 to 1527 of SEQ ID NO:19 (SEQ ID NO:21), encodes a 447 amino acid protein (Figure 18A-18C; SEQ ID NO:20).

In one embodiment of a nucleotide sequence of mouse TANGO 209 the nucleotide at position 381 is a guanine (G)(SEQ ID NO:19). In this embodiment, the amino acid at position 65 is glutamate (E)(SEQ ID NO:20). In another embodiment of a nucleotide sequence of mouse TANGO 209, the nucleotide at position 381 is a cytosine (C)(SEQ ID NO:132). In this embodiment, the amino acid at position 65 is aspartate (D)(SEQ ID NO:133). In another embodiment of a nucleotide sequence of mouse TANGO 209, the nucleotide at position 417 is an guanine (G)(SEQ ID NO:19). In this embodiment, the amino acid at position 77 is a glutamate (E)(SEQ ID NO:20). In another embodiment of a nucleotide sequence of mouse TANGO 209, the nucleotide at position 417 is a cytosine (C)(SEQ ID NO:134). In this embodiment, the amino acid at position 77

is aspartate (D)(SEQ ID NO:135). In another embodiment of a nucleotide sequence of mouse TANGO 209, the nucleotide at position 465 is a guanine (G)(SEQ ID NO:19). In this embodiment, the amino acid at position 93 is a glutamate (E)(SEQ ID NO:20). In another embodiment of a nucleotide sequence of mouse TANGO 209, the nucleotide at position 465 is a cytosine (C)(SEQ ID NO:136). In this embodiment, the amino acid at position 93 is aspartate (D)(SEQ ID NO:137).

The presence of a methionine residue at positions 324, and 398 of SEQ ID NO:20 indicate that there can be alternative forms of mouse TANGO 209 of 124 amino acids of SEQ ID NO:20 (SEQ ID NO:138), and 50 amino acids of SEQ ID NO:20 (SEQ ID NO:139), respectively.

Another embodiment of the invention includes isolated nucleic acid molecules comprising a polynucleotide having a nucleotide sequence encoding the polypeptide having the mouse TANGO 209 amino acid sequence in SEQ ID NO:20, but lacking the N-terminal methionine residue. In this embodiment, the nucleotide sequence of mouse TANGO 209, nucleotides 190 to 1527 (SEQ ID NO:140), encodes the mouse TANGO 209 amino acid sequence comprising amino acids 2-487 (SEQ ID NO:141).

The signal peptide prediction program SIGNALP (Nielsen et al. (1997) *Protein Engineering* 10:1-6) predicted that murine TANGO 209 includes a 21 amino acid signal peptide (amino acid 1 to about amino acid 21 of SEQ ID NO:20; SEQ ID NO:34) preceding the mature murine TANGO 209 protein (corresponding to about amino acid 22 to amino acid 447 of SEQ ID NO:20; SEQ ID NO:35).

Murine TANGO 209 that has not been post-translationally modified is predicted to have a molecular weight of 49.9 kDa prior to cleavage of its signal peptide and a molecular weight of 47.5 kDa subsequent to cleavage of its signal peptide.

Murine TANGO 209 includes thyroglobulin type 1 repeat domains at amino acids 109-153 of SEQ ID NO:20 (SEQ ID NO:56) and amino acids 237-281 of SEQ ID NO:20 (SEQ ID NO:57) and a Kazal-type serine protease inhibitor domain at amino acids 40-84 of SEQ ID NO:20 (SEQ ID NO:58).

In situ expression analysis of TANGO 209 expression in adult mice revealed expression in the brain (hippocampus, dentate gyrus, and frontal cortex), thymus (multifocal expression), kidney (medulla and capsule), and adrenal gland (capsule). Relatively high level, widespread, multifocal expression was observed in skeletal muscle.

Multifocal expression was observed in the diaphragm. Relatively high level expression was observed in the spleen (non-follicular). Expression was observed in the bladder, where expression was highest in muscle tissue. Expression was observed in the small intestine and colon (smooth muscle, not villi). Expression was also observed in large vessels of the liver. High level, multifocal expression was observed in the heart.

Human and murine TANGO 209 sequences exhibit considerable similarity at the protein, nucleic acid, and open reading frame levels. An alignment (made using the ALIGN software (Myers and Miller (1989) CABIOS, ver. 2.0); BLOSUM 62 scoring matrix; gap penalties -12/-4), reveals a protein identity of 94.6 %. The human and murine TANGO 209 full length cDNAs are 77.7% identical, as assessed using the same software and parameters as indicated (without the BLOSUM 62 scoring matrix). In the respective ORFs, calculated in the same fashion as the full length cDNAs, human and murine TANGO 209 are 84.4% identical. The nucleotide sequence and amino acid sequence alignments of human and murine TANGO 209 can be found in Figures 19A-19C and Figure 20, respectively.

Use of TANGO 209 Nucleic Acids, Polypeptides, and Modulators Thereof

TANGO 209 polypeptides, nucleic acids, and modulators thereof, can be used to modulate the function, morphology, proliferation and/or differentiation of cells in the tissues in which it is expressed. TANGO 209 polypeptides, nucleic acids, and modulators thereof can be used to treat disorders involving inappropriate activity of a serine protease and disorders of cellular migration, proliferation, and differentiation.

As human TANGO 209 was originally found in a LPS stimulated human primary osteoblast library, TANGO 209 nucleic acids, proteins, and modulators thereof can be used to modulate the proliferation, differentiation, and/or function of cells that form bone matrix, e.g., osteoblasts and osteoclasts, and can be used to modulate the formation of bone matrix. Thus, TANGO 209 nucleic acids, proteins, and modulators thereof can be used to treat cartilage and bone associated diseases and disorders, and can play a role in bone growth, formation, and remodeling. Examples of cartilage and bone associated diseases and disorders include e.g., bone cancer, achondroplasia, myeloma, fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma, and osteoporosis.

TANGO 209 polypeptides, nucleic acids, and modulators thereof, can be used to modulate the function, morphology, proliferation and/or differentiation of cells in the tissues in which it is expressed. Such molecules can be used to treat disorders associated with abnormal or aberrant metabolism or function of cells in the tissues in which it is expressed. Tissues in which TANGO 209 is expressed include, for example, brain, skeletal muscle, thymus, liver, adrenal gland, and the kidney.

In another example, TANGO 209 polypeptides, nucleic acids, or modulators thereof, can be used to treat disorders of the brain, such as cerebral edema, hydrocephalus, brain herniations, iatrogenic disease (due to, e.g., infection, toxins, or drugs), inflammations (e.g., bacterial and viral meningitis, encephalitis, and cerebral toxoplasmosis), cerebrovascular diseases (e.g., hypoxia, ischemia, and infarction, intracranial hemorrhage and vascular malformations, and hypertensive encephalopathy), and tumors (e.g., neuroglial tumors, neuronal tumors, tumors of pineal cells, meningeal tumors, primary and secondary lymphomas, intracranial tumors, and medulloblastoma), and to treat injury or trauma to the brain.

In another example, TANGO 209 polypeptides, nucleic acids, or modulators thereof, can be used to treat hepatic (liver) disorders, such as jaundice, hepatic failure, hereditary hyperbiliruinemias (e.g., Gilbert's syndrome, Crigler-Naijar syndromes and Dubin-Johnson and Rotor's syndromes), hepatic circulatory disorders (e.g., hepatic vein thrombosis and portal vein obstruction and thrombosis) hepatitis (e.g., chronic active hepatitis, acute viral hepatitis, and toxic and drug-induced hepatitis) cirrhosis (e.g., alcoholic cirrhosis, biliary cirrhosis, and hemochromatosis), or malignant tumors (e.g., primary carcinoma, hepatoblastoma, and angiosarcoma).

In another example, TANGO 209 polypeptides, nucleic acids, or modulators thereof, can be used to treat renal (kidney) disorders, such as glomerular diseases (e.g., acute and chronic glomerulonephritis, rapidly progressive glomerulonephritis, nephrotic syndrome, focal proliferative glomerulonephritis, glomerular lesions associated with systemic disease, such as systemic lupus erythematosus, Goodpasture's syndrome, multiple myeloma, diabetes, neoplasia, sickle cell disease, and chronic inflammatory diseases), tubular diseases (e.g., acute tubular necrosis and acute renal failure, polycystic renal diseasemedullary sponge kidney, medullary cystic disease, nephrogenic diabetes, and renal tubular acidosis), tubulointerstitial diseases (e.g., pyelonephritis, drug and toxin

induced tubulointerstitial nephritis, hypercalcemic nephropathy, and hypokalemic nephropathy) acute and rapidly progressive renal failure, chronic renal failure, nephrolithiasis, vascular diseases (e.g., hypertension and nephrosclerosis, microangiopathic hemolytic anemia, atheroembolic renal disease, diffuse cortical necrosis, and renal infarcts), or tumors (e.g., renal cell carcinoma and nephroblastoma).

In another example, TANGO 209 polypeptides, nucleic acids, or modulators thereof, can be used to treat intestinal disorders, such as ischemic bowel disease, infective enterocolitis, Crohn's disease, benign tumors, malignant tumors (e.g., argentaffinomas, lymphomas, adenocarcinomas, and sarcomas), malabsorption syndromes (e.g., celiac disease, tropical sprue, Whipple's disease, and abetalipoproteinemia), obstructive lesions, hernias, intestinal adhesions, intussusception, or volvulus.

In another example, TANGO 209 polypeptides, nucleic acids, or modulators thereof, can be used to treat disorders of the adrenal cortex, such as hypoadrenalism (e.g., primary chronic or acute adrenocortical insufficiency, and secondary adrenocortical insufficiency), hyperadrenalism (Cushing's syndrome, primary hyperaldosteronism, adrenal virilism, and adrenal hyperplasia), or neoplasia (e.g., adrenal adenoma and cortical carcinoma).

Human A236

A cDNA encoding human A236 was identified by analyzing the sequences of clones present in a human osteoblast cDNA library.

This analysis led to the identification of a clone, fthwa195d06, encoding full-length human A236. The cDNA of this clone is 1948 nucleotides long (Figure 21A-21B; SEQ ID NO:10). The 1119 nucleotide open reading frame of this cDNA, nucleotides 314-1432 of SEQ ID NO:10 (SEQ ID NO:12), encodes a 373 amino acid protein (Figure 21A-21B; SEQ ID NO:11).

In one embodiment of a nucleotide sequence of human A236 the nucleotide at position 379 is a guanine (G)(SEQ ID NO:10). In this embodiment, the amino acid at position 22 is glutamate (E)(SEQ ID NO:11). In another embodiment of a nucleotide sequence of human A236, the nucleotide at position 379 is a cytosine (C)(SEQ ID NO:142). In this embodiment, the amino acid at position 22 is aspartate (D)(SEQ ID NO:143). In another embodiment of a nucleotide sequence of mouse A236, the

nucleotide at position 397 is a guanine (G)(SEQ ID NO:10). In this embodiment, the amino acid at position 28 is a glutamate (E)(SEQ ID NO:11). In another embodiment of a nucleotide sequence of human A236, the nucleotide at position 397 is a cytosine (C)(SEQ ID NO:144). In this embodiment, the amino acid at position 28 is aspartate (D)(SEQ ID NO:145). In another embodiment of a nucleotide sequence of human A236, the nucleotide at position 400 is an adenine (A)(SEQ ID NO:10). In this embodiment, the amino acid at position 29 is a glutamate (E)(SEQ ID NO:11). In another embodiment of a nucleotide sequence of human A236, the nucleotide at position 400 is a cytosine (C)(SEQ ID NO:146). In this embodiment, the amino acid at position 29 is aspartate (D)(SEQ ID NO:147).

The presence of a methionine residue at amino acid residue positions 200, 233, and 362 of SEQ ID NO:11 indicate that there can be alternative forms of human A236 of 174 amino acids of SEQ ID NO:11 (SEQ ID NO:148), 141 amino acids of SEQ ID NO:11 (SEQ ID NO:149), and 12 amino acids of SEQ ID NO:11 (SEQ ID NO:150), respectively.

Another embodiment of the invention includes isolated nucleic acid molecules comprising a polynucleotide having a nucleotide sequence encoding the polypeptide having the human A236 amino acid sequence in SEQ ID NO:11, but lacking the N-terminal methionine residue. In this embodiment, the nucleotide sequence of human A236, nucleotides 317-1432 (SEQ ID NO:151), encodes the human A236 amino acid sequence comprising amino acids 2-373 (SEQ ID NO:152).

The signal peptide prediction program SIGNALP (Nielsen et al. (1997) *Protein Engineering* 10:1-6) predicted that human A236 includes an 18 amino acid signal peptide (amino acid 1 to about amino acid 18 of SEQ ID NO:11; SEQ ID NO:28) preceding the mature human A236 protein (corresponding to about amino acid 19 to amino acid 373 of SEQ ID NO:11; SEQ ID NO:29).

In one embodiment, human A236 has an extracellular domain which extends from about amino acid 19 to about amino acid 230 of SEQ ID NO:11 (SEQ ID NO:59), a transmembrane domain which extends from about amino acid 231 to about amino acid 255 of SEQ ID NO:11 (SEQ ID NO:60), and a cytoplasmic domain which extends from about amino acid 256 to amino acid 373 of SEQ ID NO:11 (SEQ ID NO:61). Alternatively, in another embodiment, a human A236 protein contains an extracellular domain at amino acid residues to 256 to 373 of SEQ ID NO:11 (SEQ ID NO:153), a

transmembrane domain at amino acid residues 231 to 255 of SEQ ID NO:11 (SEQ ID NO:154), and a cytoplasmic domain at amino acid residues 19 to 230 of SEQ ID NO:11 (SEQ ID NO:155).

Human A236 includes immunoglobulin domains at amino acids 28-113 of SEQ ID NO:11 (SEQ ID NO:62) and amino acids 146-210 of SEQ ID NO:11 (SEQ ID NO:63). Figure 23 depicts an alignment of the immunoglobulin domains of human A236 with a consensus immunoglobulin domain derived from a hidden Markov model (SEQ ID NO:182).

Human A236 that has not been post-translationally modified is predicted to have a molecular weight of 41.2 kDa prior to cleavage of its signal peptide and a molecular weight of 39.2 kDa subsequent to cleavage of its signal peptide.

N-glycosylation sites are present at amino acids 74-77, 197-200, and 352-355 of SEQ ID NO:11. Protein kinase C phosphorylation sites are present at amino acids 67-69, 110-112, 116-118, 296-298, 303-305, and 314-316 of SEQ ID NO:11. Casein kinase II phosphorylation sites are present at amino acids 19-22, 54-57, 157-160, 183-186, and 354-357 of SEQ ID NO:11. Tyrosine kinase phosphorylation sites are present at amino acids amino acids 102-109 and 257-264 of SEQ ID NO:11. N-myristoylation sites are present at amino acids 15-20, 146-151, 204-209, 211-216, 232-237, 240-245, 293-298, and 300-305 of SEQ ID NO:11.

Clone fthwa195d06, which encodes human A236, was deposited as Tanog 236 with the American Type Culture Collection (10801 University Boulevard, Manassas, VA 20110-2236) on May 7, 1999 and assigned Accession Number PTA-34. This deposit will be maintained under the terms of the Budapest Treaty on the International Recognition of the Deposit of Microorganisms for the Purposes of Patent Procedure. This deposit was made merely as a convenience for those of skill in the art and is not an admission that a deposit is required under 35 U.S.C. §112.

Figure 22 depicts a hydropathy plot of human A236. Relatively hydrophobic regions are above the horizontal line, and relatively hydrophilic regions are below the horizontal line. The cysteine residues (cys) are indicated by short vertical lines just below the hydropathy trace. The hydropathy plot of Figure 22 indicates the presence of a signal sequence at the amino-terminus of human A236 and a transmembrane domain within human A236, suggesting that human A236 is a transmembrane protein.

When A236 was expressed in 293T cells the cells were found to secrete a 30 kD form of A236. Briefly, 293T cells (8X105 293T cells/well) were plated and incubated in growth medium (DMEM, 10%FBS, P/S) at 37°C, 5% CO2 overnight. The cells were then transfected with an expression vector capable of expressing human A236. The transfection was performed according to the LipofectAMINE protocol (Gibco/BRL; Gaithersburg, MD) using 2 mg DNA and 10ml LipofectAMINE for each well. After the cells were transfected for 5 hrs, the culture supernatant was replaced with fresh groth medium, and the cells were incubated overnight. Next, the cells were pulse labeled as follows. The cells were washed twice with DMEM lacking methionine and cysteine. Next, 1 ml DMEM lacking methionine and cysteine and 50 mCi Trans-35S (ICN Cat#51006) was added to each well. After incubation, 150 ml samples of cell culture supernatant were collected and mixed with an equal amount of 2X SDS gel sample buffer. The samples were boiled for 5 mins and then separated by SDS PAGE.

Mouse A236

A cDNA encoding murine A236 was identified by analyzing the sequences of clones present in a mouse osteoblast cDNA library. The original mouse clone, jymuf004e01, was derived from lung.

This analysis led to the identification of a clone, jymuf004e01, encoding full-length murine A236. The cDNA of this clone is 1949 nucleotides long (Figures 24A-24B; SEQ ID NO:156). The 1119 nucleotide open reading frame of this cDNA, nucleotides 304 to 1422 of SEQ ID NO:156 (SEQ ID NO:158), encodes a 373 amino acid protein (Figure 24A-24B; SEQ ID NO:157).

In one embodiment of a nucleotide sequence of mouse A236 the nucleotide at position 366 is a guanine (G)(SEQ ID NO:156). In this embodiment, the amino acid at position 21 is glutamate (E)(SEQ ID NO:157). In another embodiment of a nucleotide sequence of mouse A236, the nucleotide at position 366 is a cytosine (C)(SEQ ID NO:159). In this embodiment, the amino acid at position 21 is aspartate (D)(SEQ ID NO:160). In another embodiment of a nucleotide sequence of mouse A236, the nucleotide at position 384 is a guanine (G)(SEQ ID NO:156). In this embodiment, the amino acid at position 27 is a glutamate (E)(SEQ ID NO:157). In another embodiment of a nucleotide sequence of mouse A236, the nucleotide at position 384 is a cytosine

(C)(SEQ ID NO:161). In this embodiment, the amino acid at position 27 is aspartate (D)(SEQ ID NO:162). In another embodiment of a nucleotide sequence of mouse A236, the nucleotide at position 387 is an adenine (A)(SEQ ID NO:156). In this embodiment, the amino acid at position 28 is a glutamate (E)(SEQ ID NO:157). In another embodiment of a nucleotide sequence of mouse A236, the nucleotide at position 387 is a cytosine (C)(SEQ ID NO:163). In this embodiment, the amino acid at position 28 is aspartate (D)(SEQ ID NO:164).

The presence of a methionine residue at amino acid residue positions 199, and 232 of SEQ ID NO:157 indicate that there can be alternative forms of mouse A236 of 175 amino acids of SEQ ID NO:157 (SEQ ID NO:165), and 142 amino acids of SEQ ID NO:157 (SEQ ID NO:166), respectively.

Another embodiment of the invention includes isolated nucleic acid molecules comprising a polynucleotide having a nucleotide sequence encoding the polypeptide having the mouse A236 amino acid sequence in SEQ ID NO:157, but lacking the N-terminal methionine residue. In this embodiment, the nucleotide sequence of mouse A236, nucleotides 317-1432 (SEQ ID NO:167), encodes the mouse A236 amino acid sequence comprising amino acids 2-373 (SEQ ID NO:168).

The signal peptide prediction program SIGNALP (Nielsen et al. (1997) *Protein Engineering* 10:1-6) predicted that murine A236 includes an 17 amino acid signal peptide (amino acid 1 to about amino acid 17 of SEQ ID NO:157; SEQ ID NO:169) preceding the mature murine A236 protein (corresponding to about amino acid 18 to amino acid 373 of SEQ ID NO:157; SEQ ID NO:170).

In one embodiment, murine A236 has an extracellular domain which extends from about amino acid 18 to about amino acid 229 of SEQ ID NO:157 (SEQ ID NO:171), a transmembrane domain which extends from about amino acid 230 to about amino acid 254 of SEQ ID NO:157 (SEQ ID NO:172), and a cytoplasmic domain which extends from about amino acid 255 to amino acid 373 of SEQ ID NO:157 (SEQ ID NO:173). Alternatively, in another embodiment, a murine A236 protein contains an extracellular domain at amino acid residues 255 to 373 of SEQ ID NO:157 (SEQ ID NO:174), a transmembrane domain at amino acid residues 230 to 254 of SEQ ID NO:157 (SEQ ID NO:172), and a cytoplasmic domain at amino acid residues 18 to 229 of SEQ ID NO:157 (SEQ ID NO:175).

Murine A236 that has not been post-translationally modified is predicted to have a molecular weight of 41.2 kDa prior to cleavage of its signal peptide and a molecular weight of 39.2 kDa subsequent to cleavage of its signal peptide.

Murine A236 includes immunoglobulin domains at amino acids 27-112 of SEQ ID NO:157 (SEQ ID NO:176) and amino acids 145-209 of SEQ ID NO:157 (SEQ ID NO:177).

A casein kinase II phosphorylation site is present at amino acids 18-21, 53-56, 182-185, and 354-357 of SEQ ID NO:157, respectively. N-myristoylation sites are present at amino acids 14-19, 145-150, 203, 208, 210, 215, 231-236, and 239-244 of SEQ ID NO:157, respectively. Protein kinase C phosphorylation sites are present at amino acids 66-68, 109-111, 115-117, 295-297, 302-304, and 313-315 of SEQ ID NO:157, respectively. A cyclic AMP phosphosite and cGMP-dependent protein kinase phosphorylation site is present at amino acids 256-259 of SEQ ID NO:157. ASN-glycosylation N glycosylation sites are present at amino acids 73-76, and 196-199 of SEQ ID NO:157, respectively. A tyrosine kinase phosphorylation site is present at amino acids 101-108 of SEQ ID NO:157, respectively.

In situ tissue screening was performed on mouse adult and embryonic tissue to analyze the expression of mouse A236 mRNA. In summary, mouse A236 mRNA expression was detected by in situ hybridization in a few adult and numerous embryonic tissues. Adult expression was detected ubiquitously in brain and more restricted in placenta, uterus, and ovary. Embryonic expression was nearly ubiquitous with higher expression in brain, mandible, and the intestinal tract. Liver noticeably lacked expression during embryogenesis.

Human and murine A236 sequences exhibit considerable similarity at the protein, nucleic acid, and open reading frame levels. An alignment (made using the ALIGN software (Myers and Miller (1989) CABIOS, ver. 2.0); BLOSUM 62 scoring matrix; gap penalties -12/-4), reveals a protein identity of 92.5%. The human and murine A236 full length cDNAs are 83.57% identical, as assessed using the same software and parameters as indicated. In the respective ORFs, calculated in the same fashion as the full length cDNAs, human and murine A236 are 87.81% identical. The nucleotide sequence (ORF) and amino acid sequence alignments of human and murine A236 can be found in Figures 25A-25B, and Figures 26A-26B, respectively.

Use of A236 Nucleic Acids, Polypeptides, and Modulators Thereof

A236 polypeptides, nucleic acids, and modulators thereof, can be used to modulate the function, morphology, proliferation and/or differentiation of cells in the tissues in which it is expressed. In addition, based on its homology to CAR (Bergelson et al. (1997) *Science* 275:1320-23) A236 may act as a entry mediator for coxsackie B viruses and adenovirus. Thus, compounds which interfere with virus binding to A236 or compounds which reduce A236 expression can be used to interfere with viral entry. A236 polypeptides, nucleic acids, and modulators thereof can be used to treat or prevent disorders associated with infection by the coxsackie B viruses and adenovirus, e.g., cardiac infection (e.g., myocarditis or dilated cardiomyopathy), central nervous system infection (e.g., non-specific ferbile illness or meningoencephalitis), pancreatic infection (e.g., acute pancreatitis), respiratory infection (pneumonia), gastrointestinal infection, or type I diabetes.

As human A236 was originally found in a LPS stimulated human primary osteoblast library, A236 nucleic acids, proteins, and modulators thereof can be used to modulate the proliferation, differentiation, and/or function of cells that form bone matrix, e.g., osteoblasts and osteoclasts, and can be used to modulate the formation of bone matrix. Thus, A236 nucleic acids, proteins, and modulators thereof can be used to treat cartilage and bone associated diseases and disorders, and can play a role in bone growth, formation, and remodeling. Examples of cartilage and bone associated diseases and disorders include e.g., bone cancer, achondroplasia, myeloma, fibrous dysplasia, scoliosis, osteoarthritis, osteosarcoma, and osteoporosis.

As the original mouse A236 clone, jymuf004e01, was derived from lung, A236 nucleic acids, proteins, and modulators thereof can be used to modulate the proliferation, activation, development, differentiation, and/or function of lung cells. Thus, A236 polypeptides, nucleic acids, or modulators thereof, can be used to treat pulmonary (lung) disorders, such as atelectasis, pulmonary congestion or edema, chronic obstructive airway disease (e.g., emphysema, chronic bronchitis, bronchial asthma, and bronchiectasis), diffuse interstitial diseases (e.g., sarcoidosis, pneumoconiosis, hypersensitivity pneumonitis, Goodpasture's syndrome, idiopathic pulmonary hemosiderosis, pulmonary alveolar proteinosis, desquamative interstitial pneumonitis, chronic interstitial pneumonia, fibrosing alveolitis, hamman-rich syndrome, pulmonary

eosinophilia, diffuse interstitial fibrosis, Wegener's granulomatosis, lymphomatoid granulomatosis, and lipid pneumonia), or tumors (e.g., bronchogenic carcinoma, bronchiolovlveolar carcinoma, bronchial carcinoid, hamartoma, and mesenchymal tumors).

A236 polypeptides, nucleic acids, and modulators thereof, can be used to modulate the function, morphology, proliferation and/or differentiation of cells in the tissues in which it is expressed. Such molecules can be used to treat disorders associated with abnormal or aberrant metabolism or function of cells in the tissues in which it is expressed. Tissues in which A236 is expressed include, for example, brain, placenta, uterus, ovaries, intestinal tract and the heart.

In another example, A236 polypeptides, nucleic acids, or modulators thereof, can be used to treat disorders of the brain, such as cerebral edema, hydrocephalus, brain herniations, iatrogenic disease (due to, e.g., infection, toxins, or drugs), inflammations (e.g., bacterial and viral meningitis, encephalitis, and cerebral toxoplasmosis), cerebrovascular diseases (e.g., hypoxia, ischemia, and infarction, intracranial hemorrhage and vascular malformations, and hypertensive encephalopathy), and tumors (e.g., neuroglial tumors, neuronal tumors, tumors of pineal cells, meningeal tumors, primary and secondary lymphomas, intracranial tumors, and medulloblastoma), and to treat injury or trauma to the brain.

In another example, A236 polypeptides, nucleic acids, or modulators thereof, can be used to treat pancreatic disorders, such as pancreatitis (e.g., acute hemorrhagic pancreatitis and chronic pancreatitis), pancreatic cysts (e.g., congenital cysts, pseudocysts, and benign or malignant neoplastic cysts), pancreatic tumors (e.g., pancreatic carcinoma and adenoma), diabetes mellitus (e.g., insulin- and non-insulin-dependent types, impaired glucose tolerance, and gestational diabetes), or islet cell tumors (e.g., insulinomas, adenomas, Zollinger-Ellison syndrome, glucagonomas, and somatostatinoma).

Because A236 is expressed in the reproductive tract, particularly in the ovaries, the A236 polypeptides, nucleic acids and/or modulators thereof can be used to modulate the function, morphology, proliferation and/or differentiation of cells in the tissues in which it is expressed.

For example, the A236 polypeptides, nucleic acids and/or modulators thereof can be used modulate the function, morphology, proliferation and/or differentiation of the

ovaries. For example, such molecules can be used to treat or modulate disorders associated with the ovaries, including, without limitation, ovarian tumors, McCune-Albright syndrome (polyostotic fibrous dysplasia). For example, the A236 polypeptides, nucleic acids and/or modulators can be used in the treatment of infertility.

The A236 polypeptides, nucleic acids and/or modulators thereof can be used to modulate the function, morphology, proliferation and/or differentiation of cells in the tissues of the reproductive tract other than the ovaries. For example, such molecules can be used to treat or modulate disorders associated with the female reproductive tract including, without limitation, uterine disorders, e.g., hyperplasia of the endometrium, uterine cancers (e.g., uterine leiomyomoma, uterine cellular leiomyoma, leiomyosarcoma of the uterus, malignant mixed mullerian Tumor of uterus, uterine Sarcoma), and dysfunctional uterine bleeding (DUB).

In another example, A236 polypeptides, nucleic acids, or modulators thereof, can be used to treat renal (kidney) disorders, such as glomerular diseases (e.g., acute and chronic glomerulonephritis, rapidly progressive glomerulonephritis, nephrotic syndrome, focal proliferative glomerulonephritis, glomerular lesions associated with systemic disease, such as systemic lupus erythematosus, Goodpasture's syndrome, multiple myeloma, diabetes, neoplasia, sickle cell disease, and chronic inflammatory diseases), tubular diseases (e.g., acute tubular necrosis and acute renal failure, polycystic renal diseasemedullary sponge kidney, medullary cystic disease, nephrogenic diabetes, and renal tubular acidosis), tubulointerstitial diseases (e.g., pyelonephritis, drug and toxin induced tubulointerstitial nephritis, hypercalcemic nephropathy, and hypokalemic nephropathy) acute and rapidly progressive renal failure, chronic renal failure, nephrolithiasis, vascular diseases (e.g., hypertension and nephrosclerosis, microangiopathic hemolytic anemia, atheroembolic renal disease, diffuse cortical necrosis, and renal infarcts), or tumors (e.g., renal cell carcinoma and nephroblastoma).

In another example, A236 polypeptides, nucleic acids, or modulators thereof, can be used to treat intestinal disorders, such as ischemic bowel disease, infective enterocolitis, Crohn's disease, benign tumors, malignant tumors (e.g., argentaffinomas, lymphomas, adenocarcinomas, and sarcomas), malabsorption syndromes (e.g., celiac disease, tropical sprue, Whipple's disease, and abetalipoproteinemia), obstructive lesions, hernias, intestinal adhesions, intussusception, or volvulus.

Tables 1 and 2 below provide summaries of TANGO 204, TANGO 206, TANGO 209, and A236 sequence information.

TABLE 1: Summary of Sequence Information for Human TANGO 204, TANGO 206, TANGO 209, and A236 and Murine TANGO 204, TANGO 206, TANGO 209, and A236.

Gene	cDNA	ORF	Polypeptide	Figure	ATCC®
TANGO 204	SEQ ID NO:1	SEQ ID NO:3	SEQ ID NO:2	Fig. 1A-1B	207192
TANGO 206	SEQ ID NO:4	SEQ ID NO:6	SEQ ID NO:5	Fig. 8A-8B	207223
TANGO 209	SEQ ID NO:7	SEQ ID NO:9	SEQ ID NO:8	Fig. 14A-14C	207223
A236 Human	SEQ ID NO:10	SEQ ID NO:12	SEQ ID NO:11	Fig. 21A-21B	PTA-34
TANGO 204	SEQ ID NO:13	SEQ ID NO:15	SEQ ID NO:14	Fig. 5A-5B	207189
TANGO 206	SEQ ID NO:16	SEQ ID NO:18	SEQ ID NO:17	Fig. 11A-11B	207221
TANGO 209	SEQ ID NO:19	SEQ ID NO:21	SEQ ID NO:20	Fig. 18A-18C	207221
Mouse A236	SEQ ID NO:156	SEQ ID NO:158	SEQ ID NO:157	Fig. 24A-24B	---

TABLE 2: Summary of Domains of Human TANGO 204, TANGO 206, TANGO 209, and A236 and Murine TANGO 204, TANGO 206, TANGO 209, and A236

Protein	Signal Sequence	Mature Protein	Extracellular Domain	Transmembrane Domain	Cytoplasmic Domain
TANGO 204 human	AA 1-20 SEQ ID NO:22	AA 21-264 SEQ ID NO:23	---	---	---
TANGO 206 human	AA 1-29 SEQ ID NO:24	AA 30-420 SEQ ID NO:25	AA 30-362 SEQ ID NO:42 AA 406-420 SEQ ID NO:46	AA 363-379 SEQ ID NO:43 AA 387-405 SEQ ID NO:45	AA 3860-386 SEQ ID NO:44
TANGO 209 human	AA 1-21 SEQ ID NO:26	AA 21-446 SEQ ID NO:27	---	---	---
A236 human	AA 1-18 SEQ ID NO:28	AA 19-373 SEQ ID NO:29	AA 19-230 SEQ ID NO:59	AA 231-255 SEQ ID NO:60	AA 256-373 SEQ ID NO:61
TANGO 204 murine	AA 1-20 SEQ ID NO:30	AA 21-264 SEQ ID NO:31	---	---	---
TANGO 206 murine	AA 1-29 SEQ ID NO:32	AA 30-420 SEQ ID NO:33	AA 30-362 SEQ ID NO:65 AA 406-420 SEQ ID NO:68	AA 363-379 SEQ ID NO:66 AA 387-405 SEQ ID NO:69	AA 380-386 SEQ ID NO:67
TANGO 209 murine	AA 1-21 SEQ ID NO:34	AA 22-447 SEQ ID NO:35	---	---	---
A236 murine	AA 1-17 SEQ ID NO:169	AA 18-373 SEQ ID NO:170	AA 18-229 SEQ ID NO:171	AA 230-254 SEQ ID NO:172	AA 255-373 SEQ ID NO:173

Deposit Information

Clone Athoc49b12, containing a cDNA molecule encoding human TANGO 206, and clone Athoc22d3, containing a cDNA molecule encoding human TANGO 209, were deposited with the American Type Culture Collection (10801 University Boulevard, Manassas, VA 20110-2236) on April 21, 1999 as Accession Number 207223. Human TANGO 206 clone Athoc49b12, deposited as EpT206, and human TANGO 209 clone Athoc22d3, deposited as EpT209, were deposited as part of a composite deposit containing five strains, each of which carries a recombinant plasmid that harbors a particular cDNA clone.

To distinguish the strains and isolate a strain harboring a particular cDNA clone, an aliquot of the mixture is streaked out on nutrient medium (e.g., LB plates) supplemented with 100 mg/ml ampicillin to isolate single colonies from which plasmid DNA is extracted using a standard minipreparation procedure. Next, a sample of the DNA minipreparation is digested with a combination of the restriction enzymes Sal I, Not I, and Sac II and the resultant products are resolved on a 0.8% agarose gel using standard DNA electrophoresis conditions. If the selected strain carries a recombinant plasmid harboring the human TANGO 206 clone, a 37 bp and 1.8 kb fragment will be generated by the digestion. If the selected strain carries a recombinant plasmid harboring the human TANGO 209 clone, a 3.1 kb fragment will be generated by the digestion.

Clone AtmMa206, containing a cDNA encoding murine TANGO 206, and clone Atmoa99h11, containing a cDNA encoding murine TANGO 209, were deposited with the American Type Culture Collection (10801 University Boulevard, Manassas, VA 20110-2236) on April 21, 1999 as Accession Number 207221. Murine TANGO 206 clone AtmMa206, deposited as EpTm206, and murine TANGO 209 clone Atmoa99h11, deposited as EpTm209, were deposited as part of a composite deposit containing five strains, each of which carries a recombinant plasmid that harbors a particular cDNA clone.

To distinguish the strains and isolate a strain harboring a particular cDNA clone, an aliquot of the mixture is streaked out on nutrient medium (e.g., LB plates) supplemented with 100 mg/ml ampicillin to isolate single colonies from which plasmid DNA is extracted using a standard minipreparation procedure. Next, a sample of the DNA minipreparation is digested with a combination of the restriction enzymes Sal I, Not

I, and Hind III and the resultant products are resolved on a 0.8% agarose gel using standard DNA electrophoresis conditions. If the selected strain carries a recombinant plasmid harboring the murine TANGO 206 clone, a 2.1 kb fragment will be generated by the digestion. If the selected strain carries a recombinant plasmid harboring the murine TANGO 209 clone, a 1.9 kb fragment and a 0.9 kb fragment will be generated by the digestion.

Clone Athu204c, which contains a cDNA encoding human TANGO 204, was deposited as fthv204c with the American Type Culture Collection (10801 University Boulevard, Manassas, VA 20110-2209) on April 2, 1999 and assigned Accession Number 207192.

Clone Atmoa043g03, which contains a cDNA encoding murine TANGO 204, was deposited as Atmoa43g3 with the American Type Culture Collection (10801 University Boulevard, Manassas, VA 20110-2209) on April 2, 1999 and assigned Accession Number 207189.

Clone fthwa195d06, which contains a cDNA encoding human A236, was deposited as Tanog 236 with the American Type Culture Collection (10801 University Boulevard, Manassas, VA 20110-2236) on May 7, 1999 and assigned Accession Number PTA-34.

Various aspects of the invention are described in further detail in the following subsections.

I. Isolated Nucleic Acid Molecules

One aspect of the invention pertains to isolated nucleic acid molecules that encode a polypeptide of the invention or a biologically active portion thereof, as well as nucleic acid molecules sufficient for use as hybridization probes to identify nucleic acid molecules encoding a polypeptide of the invention and fragments of such nucleic acid molecules suitable for use as PCR primers for the amplification or mutation of nucleic acid molecules. As used herein, the term "nucleic acid molecule" is intended to include DNA molecules (e.g., cDNA or genomic DNA) and RNA molecules (e.g., mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA.

An "isolated" nucleic acid molecule is one which is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid molecule. Preferably, an "isolated" nucleic acid molecule is free of sequences (preferably protein encoding sequences) which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated nucleic acid molecule can contain less than about 5 kB, 4 kB, 3 kB, 2 kB, 1 kB, 0.5 kB or 0.1 kB of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell from which the nucleic acid is derived. Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or substantially free of chemical precursors or other chemicals when chemically synthesized.

A nucleic acid molecule of the present invention, e.g., a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 70, 72, 74, 84, 86, 88, 90, 96, 98, 100, 109, 111, 113, 122, 124, 126, 132, 134, 136, 142, 144, 146, 156, 159, 161, or 163, or a complement thereof, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequences of SEQ ID NO:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 70, 72, 74, 84, 86, 88, 90, 96, 98, 100, 109, 111, 113, 122, 124, 126, 132, 134, 136, 142, 144, 146, 156, 159, 161, or 163 as a hybridization probe, nucleic acid molecules of the invention can be isolated using standard hybridization and cloning techniques (e.g., as described in Sambrook et al., eds., *Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

A nucleic acid molecule of the invention can be amplified using cDNA, mRNA or genomic DNA as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore, oligonucleotides corresponding to all or a portion of a nucleic acid molecule of the invention can be prepared by standard synthetic techniques, e.g., using an automated DNA synthesizer.

In another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which is a complement of the nucleotide sequence of SEQ ID NO:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 70, 72, 74, 84, 86, 88, 90, 96, 98, 100, 109, 111, 113, 122, 124, 126, 132, 134, 136, 142, 144, 146, 156, 159, 161, or 163, or a portion thereof. A nucleic acid molecule which is complementary to a given nucleotide sequence is one which is sufficiently complementary to the given nucleotide sequence that it can hybridize to the given nucleotide sequence thereby forming a stable duplex.

Moreover, a nucleic acid molecule of the invention can comprise only a portion of a nucleic acid sequence encoding a full length polypeptide of the invention for example, a fragment which can be used as a probe or primer or a fragment encoding a biologically active portion of a polypeptide of the invention. The nucleotide sequence determined from the cloning one gene allows for the generation of probes and primers designed for use in identifying and/or cloning homologues in other cell types, e.g., from other tissues, as well as homologues from other mammals. The probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, preferably about 25, more preferably about 50, 75, 100, 125, 150, 175, 200, 250, 300, 350 or 400 consecutive nucleotides of the sense or anti-sense sequence of SEQ ID NO:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 70, 72, 74, 84, 86, 88, 90, 96, 98, 100, 109, 111, 113, 122, 124, 126, 132, 134, 136, 142, 144, 146, 156, 159, 161, or 163, or of a naturally occurring mutant of SEQ ID NO:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 70, 72, 74, 84, 86, 88, 90, 96, 98, 100, 109, 111, 113, 122, 124, 126, 132, 134, 136, 142, 144, 146, 156, 159, 161, or 163.

Probes based on the sequence of a nucleic acid molecule of the invention can be used to detect transcripts or genomic sequences encoding the same protein molecule encoded by a selected nucleic acid molecule. The probe comprises a label group attached thereto, e.g., a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as part of a diagnostic test kit for identifying cells or tissues which mis-express the protein, such as by measuring levels of a nucleic acid molecule encoding the protein in a sample of cells from a subject, e.g., detecting mRNA levels or determining whether a gene encoding the protein has been mutated or deleted.

A nucleic acid fragment encoding a a biologically active portion of a polypeptide of the invention can be prepared by isolating a portion of any of SEQ ID NO:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 70, 72, 74, 84, 86, 88, 90, 96, 98, 100, 109, 111, 113, 122, 124, 126, 132, 134, 136, 142, 144, 146, 156, 159, 161, or 163, expressing the encoded portion of the polypeptide protein (e.g., by recombinant expression *in vitro*) and assessing the activity of the encoded portion of the polypeptide.

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequence of SEQ ID NO:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 70, 72, 74, 84, 86, 88, 90, 96, 98, 100, 109, 111, 113, 122, 124, 126, 132, 134, 136, 142, 144, 146, 156, 159, 161, or 163, due to degeneracy of the genetic code and thus encode the same protein as that encoded by the nucleotide sequence of SEQ ID NO:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 70, 72, 74, 84, 86, 88, 90, 96, 98, 100, 109, 111, 113, 122, 124, 126, 132, 134, 136, 142, 144, 146, 156, 159, 161, or 163.

In addition to the nucleotide sequences of SEQ ID NO:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 70, 72, 74, 84, 86, 88, 90, 96, 98, 100, 109, 111, 113, 122, 124, 126, 132, 134, 136, 142, 144, 146, 156, 159, 161, or 163, it will be appreciated by those skilled in the art that DNA sequence polymorphisms that lead to changes in the amino acid sequence may exist within a population (e.g., the human population). Such genetic polymorphisms may exist among individuals within a population due to natural allelic variation. An allele is one of a group of genes which occur alternatively at a given genetic locus. As used herein, the phrase "allelic variant" refers to a nucleotide sequence which occurs at a given locus or to a polypeptide encoded by the nucleotide sequence. given locus or to a polypeptide encoded by the nucleotide sequence. For example, human TANGO 204 has been mapped on radiation hybrid panels to the long arm of chromosome 8q, in the region , between flanking markers D1Mit430 and D1Mit119, and therefore, human TANGO 204 family members can include nucleotide sequence polymorphisms (*e.g.*, nucleotide sequences that vary from SEQ ID NO:1) that map to this chromosome 8 region (*i.e.*, between markers D1Mit430 and D1Mit119). For example, the human gene for TANGO 209 has been mapped on radiation hybrid panels to the long arm of chromosome 6, in the region q26-27, between flanking markers ATA22G07 and WI-9405, and therefore, human TANGO 209 family members can include nucleotide sequence polymorphisms (*e.g.*, nucleotide sequences that vary from SEQ ID NO:7) that map to this

chromosome 6 region (*i.e.*, between markers ATA22G07 and WI-9405). As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame encoding a polypeptide of the invention. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of a given gene. Alternative alleles can be identified by sequencing the gene of interest in a number of different individuals. This can be readily carried out by using hybridization probes to identify the same genetic locus in a variety of individuals. Any and all such nucleotide variations and resulting amino acid polymorphisms or variations that are the result of natural allelic variation and that do not alter the functional activity are intended to be within the scope of the invention.

Moreover, nucleic acid molecules encoding proteins of the invention from other species (homologues), which have a nucleotide sequence which differs from that of the human protein described herein are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and homologues of a cDNA of the invention can be isolated based on their identity to the human nucleic acid molecule disclosed herein using the human cDNAs, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions. For example, a cDNA encoding a soluble form of a membrane-bound protein of the invention isolated based on its hybridization to a nucleic acid molecule encoding all or part of the membrane-bound form. Likewise, a cDNA encoding a membrane-bound form can be isolated based on its hybridization to a nucleic acid molecule encoding all or part of the soluble form.

Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 300 (325, 350, 375, 400, 425, 450, 500, 550, 600, 650, 700, 800, 900, 1000, or 1200, 1400, 1600, 1800, 2000, 2200, 2400, 2600, 2800, 3000, or 3200) nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence, preferably the coding sequence, of SEQ ID NO:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 70, 72, 74, 84, 86, 88, 90, 96, 98, 100, 109, 111, 113, 122, 124, 126, 132, 134, 136, 142, 144, 146, 156, 159, 161, or 163, or a complement thereof.

As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at

least 60% (65%, 70%, preferably 75%) identical to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. A preferred, non-limiting example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45EC, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65EC. Preferably, an isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to the sequence of SEQ ID NO:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 70, 72, 74, 84, 86, 88, 90, 96, 98, 100, 109, 111, 113, 122, 124, 126, 132, 134, 136, 142, 144, 146, 156, 159, 161, or 163, or a complement thereof, corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein).

In addition to naturally-occurring allelic variants of a nucleic acid molecule of the invention sequence that may exist in the population, the skilled artisan will further appreciate that changes can be introduced by mutation thereby leading to changes in the amino acid sequence of the encoded protein, without altering the biological activity of the protein. For example, one can make nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid residues. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequence without altering the biological activity, whereas an "essential" amino acid residue is required for biological activity. For example, amino acid residues that are not conserved or only semi-conserved among homologues of various species may be non-essential for activity and thus would be likely targets for alteration. Alternatively, amino acid residues that are conserved among the homologues of various species (e.g., murine and human) may be essential for activity and thus would not be likely targets for alteration.

Accordingly, another aspect of the invention pertains to nucleic acid molecules encoding a polypeptide of the invention that contain changes in amino acid residues that are not essential for activity. Such polypeptides differ in amino acid sequence from SEQ ID Nos. 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 12, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164, yet retain biological activity. In one embodiment, the isolated nucleic acid molecule includes a nucleotide sequence

encoding a protein that includes an amino acid sequence that is at least about 45% identical, 65%, 75%, 85%, 95%, or 98% identical to the amino acid sequence of SEQ ID Nos. 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 12, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164.

An isolated nucleic acid molecule encoding a variant protein can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NO:1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 70, 72, 74, 84, 86, 88, 90, 96, 98, 100, 109, 111, 113, 122, 124, 126, 132, 134, 136, 142, 144, 146, 156, 159, 161, or 163, such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein. Mutations can be introduced by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined in the art. These families include amino acids with basic side chains (e.g., lysine, arginine, histidine), acidic side chains (e.g., aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine, tryptophan, histidine). Alternatively, mutations can be introduced randomly along all or part of the coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for biological activity to identify mutants that retain activity. Following mutagenesis, the encoded protein can be expressed recombinantly and the activity of the protein can be determined.

In one embodiment, a mutant polypeptide that is a variant of a polypeptide of the invention can be assayed for: (1) the ability to form protein:protein interactions with proteins in a signaling pathway of the polypeptide of the invention; (2) the ability to bind a ligand of the polypeptide of the invention; or (3) the ability to bind to an intracellular target protein of the polypeptide of the invention. In yet another embodiment, the mutant polypeptide can be assayed for the ability to modulate cellular proliferation, cellular migration or chemotaxis, or cellular differentiation.

The present invention encompasses antisense nucleic acid molecules, i.e., molecules which are complementary to a sense nucleic acid encoding a polypeptide of the invention, e.g., complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence. Accordingly, an antisense nucleic acid can hydrogen bond to a sense nucleic acid. The antisense nucleic acid can be complementary to an entire coding strand, or to only a portion thereof, e.g., all or part of the protein coding region (or open reading frame). An antisense nucleic acid molecule can be antisense to all or part of a non-coding region of the coding strand of a nucleotide sequence encoding a polypeptide of the invention. The non-coding regions ("5' and 3' untranslated regions") are the 5' and 3' sequences which flank the coding region and are not translated into amino acids.

An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (e.g., an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, e.g., phosphorothioate derivatives and acridine substituted nucleotides can be used. Examples of modified nucleotides which can be used to generate the antisense nucleic acid include 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-(carboxyhydroxylmethyl) uracil, 5-carboxymethylaminomethyl-2-thiouridine, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 5-methoxyuracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (v), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 2-thiouracil, 4-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (v), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)w, and 2,6-diaminopurine. Alternatively, the

antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (i.e., RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a selected polypeptide of the invention to thereby inhibit expression, e.g., by inhibiting transcription and/or translation. The hybridization can be by conventional nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule which binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site. Alternatively, antisense nucleic acid molecules can be modified to target selected cells and then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface, e.g., by linking the antisense nucleic acid molecules to peptides or antibodies which bind to cell surface receptors or antigens. The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve sufficient intracellular concentrations of the antisense molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

An antisense nucleic acid molecule of the invention can be an "-anomeric nucleic acid molecule. An alpha-anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual beta-units, the strands run parallel to each other (Gaultier et al. (1987) *Nucleic Acids Res.* 15:6625-6641). The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (Inoue et al. (1987) *Nucleic Acids Res.* 15:6131-6148) or a chimeric RNA-DNA analogue (Inoue et al. (1987) *FEBS Lett.* 215:327-330).

The invention also encompasses ribozymes. Ribozymes are catalytic RNA molecules with ribonuclease activity which are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a complementary region. Thus,

ribozymes (e.g., hammerhead ribozymes (described in Haselhoff and Gerlach (1988) *Nature* 334:585-591)) can be used to catalytically cleave mRNA transcripts to thereby inhibit translation of the protein encoded by the mRNA. A ribozyme having specificity for a nucleic acid molecule encoding a polypeptide of the invention can be designed based upon the nucleotide sequence of a cDNA disclosed herein. For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a Cech et al. U.S. Patent No. 4,987,071; and Cech et al. U.S. Patent No. 5,116,742. Alternatively, an mRNA encoding a polypeptide of the invention can be used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. *See, e.g.*, Bartel and Szostak (1993) *Science* 261:1411-1418.

The invention also encompasses nucleic acid molecules which form triple helical structures. For example, expression of a polypeptide of the invention can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the gene encoding the polypeptide (e.g., the promoter and/or enhancer) to form triple helical structures that prevent transcription of the gene in target cells. *See generally* Helene (1991) *Anticancer Drug Des.* 6(6):569-84; Helene (1992) *Ann. N.Y. Acad. Sci.* 660:27-36; and Maher (1992) *Bioassays* 14(12):807-15.

In various embodiments, the nucleic acid molecules of the invention can be modified at the base moiety, sugar moiety or phosphate backbone to improve, e.g., the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can be modified to generate peptide nucleic acids (*see* Hyrup et al. (1996) *Bioorganic & Medicinal Chemistry* 4(1): 5-23). As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics, e.g., DNA mimics, in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleobases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and RNA under conditions of low ionic strength. The synthesis of PNA oligomers can be performed using standard solid phase peptide synthesis protocols as described in Hyrup et al. (1996), *supra*; Perry-O'Keefe et al. (1996) *Proc. Natl. Acad. Sci. USA* 93: 14670-675.

PNAs can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of

gene expression by, e.g., inducing transcription or translation arrest or inhibiting replication. PNAs can also be used, e.g., in the analysis of single base pair mutations in a gene by, e.g., PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, e.g., S1 nucleases (Hyrup (1996), *supra*; or as probes or primers for DNA sequence and hybridization (Hyrup (1996), *supra*; Perry-O'Keefe et al. (1996) *Proc. Natl. Acad. Sci. USA* 93: 14670-675).

In another embodiment, PNAs can be modified, e.g., to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras can be generated which may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes, e.g., RNAse H and DNA polymerases, to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleobases, and orientation (Hyrup (1996), *supra*). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup (1996), *supra,* and Finn et al. (1996) *Nucleic Acids Res.* 24(17):3357-63. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry and modified nucleoside analogs. Compounds such as 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite can be used as a link between the PNA and the 5' end of DNA (Mag et al. (1989) *Nucleic Acids Res.* 17:5973-88). PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment (Finn et al. (1996) *Nucleic Acids Res.* 24(17):3357-63). Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment (Peterser et al. (1975) *Bioorganic Med. Chem. Lett.* 5:1119-11124).

In other embodiments, the oligonucleotide may include other appended groups such as peptides (e.g., for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (*see, e.g.*, Letsinger et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:6553-6556; Lemaitre et al. (1987) *Proc. Natl. Acad. Sci. USA* 84:648-652; PCT Publication No. W0 88/09810) or the blood-brain barrier (*see, e.g.*, PCT Publication No. W0 89/10134). In addition, oligonucleotides can be modified with hybridization-triggered cleavage agents (*see, e.g.,* Krol et al. (1988) *Bio/Techniques* 6:958-976) or intercalating

agents (*see, e.g.*, Zon (1988) *Pharm. Res.* 5:539-549). To this end, the oligonucleotide may be conjugated to another molecule, e.g., a peptide, hybridization triggered cross-linking agent, transport agent, hybridization-triggered cleavage agent, etc.

II. Isolated Proteins and Antibodies

One aspect of the invention pertains to isolated proteins, and biologically active portions thereof, as well as polypeptide fragments suitable for use as immunogens to raise antibodies directed against a polypeptide of the invention. In one embodiment, the native polypeptide can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, polypeptides of the invention are produced by recombinant DNA techniques. Alternative to recombinant expression, a polypeptide of the invention can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" protein or biologically active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the protein is derived, or substantially free of chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of protein in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly produced. Thus, protein that is substantially free of cellular material includes preparations of protein having less than about 30%, 20%, 10%, or 5% (by dry weight) of heterologous protein (also referred to herein as a "contaminating protein"). When the protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, i.e., culture medium represents less than about 20%, 10%, or 5% of the volume of the protein preparation. When the protein is produced by chemical synthesis, it is preferably substantially free of chemical precursors or other chemicals, i.e., it is separated from chemical precursors or other chemicals which are involved in the synthesis of the protein. Accordingly such preparations of the protein have less than about 30%, 20%, 10%, 5% (by dry weight) of chemical precursors or compounds other than the polypeptide of interest.

Biologically active portions of a polypeptide of the invention include polypeptides comprising amino acid sequences sufficiently identical to or derived from

the amino acid sequence of the protein (e.g., the amino acid sequence shown in any of SEQ ID Nos. 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 12, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164), which include fewer amino acids than the full length protein, and exhibit at least one activity of the corresponding full-length protein. Typically, biologically active portions comprise a domain or motif with at least one activity of the corresponding protein. A biologically active portion of a protein of the invention can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acids in length. Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of the native form of a polypeptide of the invention.

Preferred polypeptides have the amino acid sequence of SEQ ID Nos. 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 12, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164. Other useful proteins are substantially identical (e.g., at least about 45%, preferably 55%, 65%, 75%, 85%, 95%, or 99%) to any of SEQ ID Nos. 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 12, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164 and retain the functional activity of the protein of the corresponding naturally-occurring protein yet differ in amino acid sequence due to natural allelic variation or mutagenesis.

To determine the percent identity of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences (i.e., % identity = # of identical positions/total # of positions (e.g., overlapping positions) x 100). In one embodiment the two sequences are the same length.

The determination of percent identity between two sequences can be accomplished using a mathematical algorithm. A preferred, non-limiting example of a

mathematical algorithm utilized for the comparison of two sequences is the algorithm of Karlin and Altschul (1990) *Proc. Natl. Acad. Sci. USA* 87:2264-2268, modified as in Karlin and Altschul (1993) *Proc. Natl. Acad. Sci. USA* 90:5873-5877. Such an algorithm is incorporated into the NBLAST and XBLAST programs of Altschul, et al. (1990) *J. Mol. Biol.* 215:403-410. BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to a nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to a protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al. (1997) *Nucleic Acids Res.* 25:3389-3402. Alternatively, PSI-Blast can be used to perform an iterated search which detects distant relationships between molecules (*Id.*). When utilizing BLAST, Gapped BLAST, and PSI-Blast programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. *See* http://www.ncbi.nlm.nih.gov. Another preferred, non-limiting example of a mathematical algorithm utilized for the comparison of sequences is the algorithm of Myers and Miller, (1988) *CABIOS* 4:11-17. Such an algorithm is incorporated into the ALIGN program (version 2.0) which is part of the GCG sequence alignment software package. When utilizing the ALIGN program for comparing amino acid sequences, a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4 can be used.

The percent identity between two sequences can be determined using techniques similar to those described above, with or without allowing gaps. In calculating percent identity, only exact matches are counted.

The invention also provides chimeric or fusion proteins. As used herein, a "chimeric protein" or "fusion protein" comprises all or part (preferably biologically active) of a polypeptide of the invention operably linked to a heterologous polypeptide (i.e., a polypeptide other than the same polypeptide of the invention). Within the fusion protein, the term "operably linked" is intended to indicate that the polypeptide of the invention and the heterologous polypeptide are fused in-frame to each other. The heterologous polypeptide can be fused to the N-terminus or C-terminus of the polypeptide of the invention.

One useful fusion protein is a GST fusion protein in which the polypeptide of the invention is fused to the C-terminus of GST sequences. Such fusion proteins can facilitate the purification of a recombinant polypeptide of the invention.

In another embodiment, the fusion protein contains a heterologous signal sequence at its N-terminus. For example, the native signal sequence of a polypeptide of the invention can be removed and replaced with a signal sequence from another protein. For example, the gp67 secretory sequence of the baculovirus envelope protein can be used as a heterologous signal sequence (*Current Protocols in Molecular Biology*, Ausubel et al., eds., John Wiley & Sons, 1992). Other examples of eukaryotic heterologous signal sequences include the secretory sequences of melittin and human placental alkaline phosphatase (Stratagene; La Jolla, California). In yet another example, useful prokaryotic heterologous signal sequences include the phoA secretory signal (Sambrook et al., *supra*) and the protein A secretory signal (Pharmacia Biotech; Piscataway, New Jersey).

In yet another embodiment, the fusion protein is an immunoglobulin fusion protein in which all or part of a polypeptide of the invention is fused to sequences derived from a member of the immunoglobulin protein family. The immunoglobulin fusion proteins of the invention can be incorporated into pharmaceutical compositions and administered to a subject to inhibit an interaction between a ligand (soluble or membrane-bound) and a protein on the surface of a cell (receptor), to thereby suppress signal transduction *in vivo*. The immunoglobulin fusion protein can be used to affect the bioavailability of a cognate ligand of a polypeptide of the invention. Inhibition of ligand/receptor interaction may be useful therapeutically, both for treating proliferative and differentiative disorders and for modulating (e.g., promoting or inhibiting) cell survival. Moreover, the immunoglobulin fusion proteins of the invention can be used as immunogens to produce antibodies directed against a polypeptide of the invention in a subject, to purify ligands and in screening assays to identify molecules which inhibit the interaction of receptors with ligands.

Chimeric and fusion proteins of the invention can be produced by standard recombinant DNA techniques. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene

fragments which can subsequently be annealed and reamplified to generate a chimeric gene sequence (*see, e.g.*, Ausubel et al., *supra*). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST polypeptide). A nucleic acid encoding a polypeptide of the invention can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the polypeptide of the invention.

A signal sequence of a polypeptide of the invention (SEQ ID NO:22, 24, 26, 28, 30, 32, 34, or 169) can be used to facilitate secretion and isolation of the secreted protein or other proteins of interest. Signal sequences are typically characterized by a core of hydrophobic amino acids which are generally cleaved from the mature protein during secretion in one or more cleavage events. Such signal peptides contain processing sites that allow cleavage of the signal sequence from the mature proteins as they pass through the secretory pathway. Thus, the invention pertains to the described polypeptides having a signal sequence, as well as to the signal sequence itself and to the polypeptide in the absence of the signal sequence (i.e., the cleavage products). In one embodiment, a nucleic acid sequence encoding a signal sequence of the invention can be operably linked in an expression vector to a protein of interest, such as a protein which is ordinarily not secreted or is otherwise difficult to isolate. The signal sequence directs secretion of the protein, such as from a eukaryotic host into which the expression vector is transformed, and the signal sequence is subsequently or concurrently cleaved. The protein can then be readily purified from the extracellular medium by art recognized methods. Alternatively, the signal sequence can be linked to the protein of interest using a sequence which facilitates purification, such as with a GST domain.

In another embodiment, the signal sequences of the present invention can be used to identify regulatory sequences, e.g., promoters, enhancers, repressors. Since signal sequences are the most amino-terminal sequences of a peptide, it is expected that the nucleic acids which flank the signal sequence on its amino-terminal side will be regulatory sequences which affect transcription. Thus, a nucleotide sequence which encodes all or a portion of a signal sequence can be used as a probe to identify and isolate signal sequences and their flanking regions, and these flanking regions can be studied to identify regulatory elements therein.

The present invention also pertains to variants of the polypeptides of the invention. Such variants have an altered amino acid sequence which can function as either agonists (mimetics) or as antagonists. Variants can be generated by mutagenesis, e.g., discrete point mutation or truncation. An agonist can retain substantially the same, or a subset, of the biological activities of the naturally occurring form of the protein. An antagonist of a protein can inhibit one or more of the activities of the naturally occurring form of the protein by, for example, competitively binding to a downstream or upstream member of a cellular signaling cascade which includes the protein of interest. Thus, specific biological effects can be elicited by treatment with a variant of limited function. Treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein can have fewer side effects in a subject relative to treatment with the naturally occurring form of the protein.

Variants of a protein of the invention which function as either agonists (mimetics) or as antagonists can be identified by screening combinatorial libraries of mutants, e.g., truncation mutants, of the protein of the invention for agonist or antagonist activity. In one embodiment, a variegated library of variants is generated by combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential protein sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (e.g., for phage display). There are a variety of methods which can be used to produce libraries of potential variants of the polypeptides of the invention from a degenerate oligonucleotide sequence. Methods for synthesizing degenerate oligonucleotides are known in the art (*see, e.g.*, Narang (1983) *Tetrahedron* 39:3; Itakura et al. (1984) *Annu. Rev. Biochem.* 53:323; Itakura et al. (1984) *Science* 198:1056; Ike et al. (1983) *Nucleic Acid Res.* 11:477).

In addition, libraries of fragments of the coding sequence of a polypeptide of the invention can be used to generate a variegated population of polypeptides for screening and subsequent selection of variants. For example, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of the coding sequence of interest with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form

double stranded DNA which can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal and internal fragments of various sizes of the protein of interest.

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. The most widely used techniques, which are amenable to high through-put analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify variants of a protein of the invention (Arkin and Yourvan (1992) *Proc. Natl. Acad. Sci. USA* 89:7811-7815; Delgrave et al. (1993) *Protein Engineering* 6(3):327-331).

An isolated polypeptide of the invention, or a fragment thereof, can be used as an immunogen to generate antibodies using standard techniques for polyclonal and monoclonal antibody preparation. The full-length polypeptide or protein can be used or, alternatively, the invention provides antigenic peptide fragments for use as immunogens. The antigenic peptide of a protein of the invention comprises at least 8 (preferably 10, 15, 20, or 30) amino acid residues of the amino acid sequence of SEQ ID Nos. 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 12, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164, and encompasses an epitope of the protein such that an antibody raised against the peptide forms a specific immune complex with the protein.

Preferred epitopes encompassed by the antigenic peptide are regions that are located on the surface of the protein, e.g., hydrophilic regions. Figures 2, 9, 15, 22, and 27 are hydropathy plots of the proteins of the invention. These plots or similar analyses can be used to identify hydrophilic regions.

An immunogen typically is used to prepare antibodies by immunizing a suitable subject, (e.g., rabbit, goat, mouse or other mammal). An appropriate

immunogenic preparation can contain, for example, recombinantly expressed or chemically synthesized polypeptide. The preparation can further include an adjuvant, such as Freund's complete or incomplete adjuvant, or similar immunostimulatory agent.

Accordingly, another aspect of the invention pertains to antibodies directed against a polypeptide of the invention. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, i.e., molecules that contain an antigen binding site which specifically binds an antigen, such as a polypeptide of the invention *e.g.*, an epitope of a polypeptide of the invention. A molecule which specifically binds to a given polypeptide of the invention is a molecule which binds the polypeptide, but does not substantially bind other molecules in a sample, e.g., a biological sample, which naturally contains the polypeptide. Examples of immunologically active portions of immunoglobulin molecules include F(ab) and $F(ab')_2$ fragments which can be generated by treating the antibody with an enzyme such as pepsin. The invention provides polyclonal and monoclonal antibodies. The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope.

Polyclonal antibodies can be prepared as described above by immunizing a suitable subject with a polypeptide of the invention as an immunogen. Preferred polyclonal antibody compositions are ones that have been selected for antibodies directed against a polypeptide or polypeptides of the invention. Particularly preferred polyclonal antibody preparations are ones that contain only antibodies directed against a polypeptide or polypeptides of the invention. Particularly preferred immunogen compositions are those that contain no other human proteins such as, for example, immunogen compositions made using a non-human host cell for recombinant expression of a polypeptide of the invention. In such a manner, the only human epitope or epitopes recognized by the resulting antibody compositions raised against this immunogen will be present as part of a polypeptide or polypeptides of the invention.

The antibody titer in the immunized subject can be monitored over time by standard techniques, such as with an enzyme linked immunosorbent assay (ELISA) using immobilized polypeptide. If desired, the antibody molecules can be isolated from the mammal (e.g., from the blood) and further purified by well-known techniques, such as

protein A chromatography to obtain the IgG fraction. Alternatively, antibodies specific for a protein or polypeptide of the invention can be selected for (*e.g.*, partially purified) or purified by, *e.g.*, affinity chromatography. For example, a recombinantly expressed and purified (or partially purified) protein of the invention is produced as described herein, and covalently or non-covalently coupled to a solid support such as, for example, a chromatography column. The column can then be used to affinity purify antibodies specific for the proteins of the invention from a sample containing antibodies directed against a large number of different epitopes, thereby generating a substantially purified antibody composition, *i.e.*, one that is substantially free of contaminating antibodies. By a substantially purified antibody composition is meant, in this context, that the antibody sample contains at most only 30% (by dry weight) of contaminating antibodies directed against epitopes other than those on the desired protein or polypeptide of the invention, and preferably at most 20%, yet more preferably at most 10%, and most preferably at most 5% (by dry weight) of the sample is contaminating antibodies. A purified antibody composition means that at least 99% of the antibodies in the composition are directed against the desired protein or polypeptide of the invention.

At an appropriate time after immunization, e.g., when the specific antibody titers are highest, antibody-producing cells can be obtained from the subject and used to prepare monoclonal antibodies by standard techniques, such as the hybridoma technique originally described by Kohler and Milstein (1975) *Nature* 256:495-497, the human B cell hybridoma technique (Kozbor et al. (1983) *Immunol. Today* 4:72), the EBV-hybridoma technique (Cole et al. (1985), *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, Inc., pp. 77-96) or trioma techniques. The technology for producing hybridomas is well known (*see generally Current Protocols in Immunology* (1994) Coligan et al. (eds.) John Wiley & Sons, Inc., New York, NY). Hybridoma cells producing a monoclonal antibody of the invention are detected by screening the hybridoma culture supernatants for antibodies that bind the polypeptide of interest, e.g., using a standard ELISA assay.

Alternative to preparing monoclonal antibody-secreting hybridomas, a monoclonal antibody directed against a polypeptide of the invention can be identified and isolated by screening a recombinant combinatorial immunoglobulin library (e.g., an antibody phage display library) with the polypeptide of interest. Kits for generating and screening phage display libraries are commercially available (e.g., the Pharmacia

Recombinant Phage Antibody System, Catalog No. 27-9400-01; and the Stratagene *SurfZAPJ Phage Display Kit*, Catalog No. 240612). Additionally, examples of methods and reagents particularly amenable for use in generating and screening antibody display library can be found in, for example, U.S. Patent No. 5,223,409; PCT Publication No. WO 92/18619; PCT Publication No. WO 91/17271; PCT Publication No. WO 92/20791; PCT Publication No. WO 92/15679; PCT Publication No. WO 93/01288; PCT Publication No. WO 92/01047; PCT Publication No. WO 92/09690; PCT Publication No. WO 90/02809; Fuchs et al. (1991) *Bio/Technology* 9:1370-1372; Hay et al. (1992) *Hum. Antibod. Hybridomas* 3:81-85; Huse et al. (1989) *Science* 246:1275-1281; Griffiths et al. (1993) *EMBO J.* 12:725-734.

Additionally, recombinant antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the invention. A chimeric antibody is a molecule in which different portions are derived from different animal species, such as those having a variable region derived from a murine mAb and a human immunoglobulin constant region. (See, *e.g.*, Cabilly et al., U.S. Patent No. 4,816,567; and Boss et al., U.S. Patent No. 4,816,397, which are incorporated herein by reference in their entirety.) Humanized antibodies are antibody molecules from non-human species having one or more complementarily determining regions (CDRs) from the non-human species and a framework region from a human immunoglobulin molecule. (See, *e.g.*, Queen, U.S. Patent No. 5,585,089, which is incorporated herein by reference in its entirety.) Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art, for example using methods described in PCT Publication No. WO 87/02671; European Patent Application 184,187; European Patent Application 171,496; European Patent Application 173,494; PCT Publication No. WO 86/01533; U.S. Patent No. 4,816,567; European Patent Application 125,023; Better et al. (1988) *Science* 240:1041-1043; Liu et al. (1987) *Proc. Natl. Acad. Sci. USA* 84:3439-3443; Liu et al. (1987) *J. Immunol.* 139:3521-3526; Sun et al. (1987) *Proc. Natl. Acad. Sci. USA* 84:214-218; Nishimura et al. (1987) *Canc. Res.* 47:999-1005; Wood et al. (1985) *Nature* 314:446-449; and Shaw et al. (1988) *J. Natl. Cancer Inst.* 80:1553-1559); Morrison (1985) *Science* 229:1202-1207; Oi et al. (1986) *Bio/Techniques*

4:214; U.S. Patent 5,225,539; Jones et al. (1986) *Nature* 321:552-525; Verhoeyan et al. (1988) *Science* 239:1534; and Beidler et al. (1988) *J. Immunol.* 141:4053-4060.

Completely human antibodies are particularly desirable for therapeutic treatment of human patients. Such antibodies can be produced, for example, using transgenic mice which are incapable of expressing endogenous immunoglobulin heavy and light chains genes, but which can express human heavy and light chain genes. The transgenic mice are immunized in the normal fashion with a selected antigen, e.g., all or a portion of a polypeptide of the invention. Monoclonal antibodies directed against the antigen can be obtained using conventional hybridoma technology. The human immunoglobulin transgenes harbored by the transgenic mice rearrange during B cell differentiation, and subsequently undergo class switching and somatic mutation. Thus, using such a technique, it is possible to produce therapeutically useful IgG, IgA and IgE antibodies. For an overview of this technology for producing human antibodies, see Lonberg and Huszar (1995, *Int. Rev. Immunol.* 13:65-93). For a detailed discussion of this technology for producing human antibodies and human monoclonal antibodies and protocols for producing such antibodies, *see, e.g.*, U.S. Patent 5,625,126; U.S. Patent 5,633,425; U.S. Patent 5,569,825; U.S. Patent 5,661,016; and U.S. Patent 5,545,806. In addition, companies such as Abgenix, Inc. (Freemont, CA), can be engaged to provide human antibodies directed against a selected antigen using technology similar to that described above.

Completely human antibodies which recognize a selected epitope can be generated using a technique referred to as "guided selection." In this approach a selected non-human monoclonal antibody, e.g., a murine antibody, is used to guide the selection of a completely human antibody recognizing the same epitope. (Jespers et al. (1994) *Bio/technology* 12:899-903).

An antibody directed against a polypeptide of the invention (e.g., monoclonal antibody) can be used to isolate the polypeptide by standard techniques, such as affinity chromatography or immunoprecipitation. Moreover, such an antibody can be used to detect the protein (e.g., in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the polypeptide. The antibodies can also be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, e.g., to, for example, determine the efficacy of a given treatment regimen. Detection can

be facilitated by coupling the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, 8-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

Further, an antibody (or fragment thereof) can be conjugated to a therapeutic moiety such as a cytotoxin, a therapeutic agent or a radioactive metal ion. A cytotoxin or cytotoxic agent includes any agent that is detrimental to cells. Examples include taxol, cytochalasin B, gramicidin D, ethidium bromide, emetine, mitomycin, etoposide, tenoposide, vincristine, vinblastine, colchicin, doxorubicin, daunorubicin, dihydroxy anthracin dione, mitoxantrone, mithramycin, actinomycin D, 1-dehydrotestosterone, glucocorticoids, procaine, tetracaine, lidocaine, propranolol, and puromycin and analogs or homologs thereof. Therapeutic agents include, but are not limited to, antimetabolites (*e.g.*, methotrexate, 6-mercaptopurine, 6-thioguanine, cytarabine, 5-fluorouracil decarbazine), alkylating agents (*e.g.*, mechlorethamine, thioepa chlorambucil, melphalan, carmustine (BSNU) and lomustine (CCNU), cyclothosphamide, busulfan, dibromomannitol, streptozotocin, mitomycin C, and cis-dichlorodiamine platinum (II) (DDP) cisplatin), anthracyclines (*e.g.*, daunorubicin (formerly daunomycin) and doxorubicin), antibiotics (*e.g.*, dactinomycin (formerly actinomycin), bleomycin, mithramycin, and anthramycin (AMC)), and anti-mitotic agents (*e.g.*, vincristine and vinblastine).

The conjugates of the invention can be used for modifying a given biological response, the drug moiety is not to be construed as limited to classical chemical therapeutic agents. For example, the drug moiety may be a protein or polypeptide possessing a desired biological activity. Such proteins may include, for example, a toxin such as abrin, ricin A, pseudomonas exotoxin, or diphtheria toxin; a protein such as tumor necrosis factor, .alpha.-interferon, .beta.-interferon, nerve growth factor, platelet derived

growth factor, tissue plasminogen activator; or, biological response modifiers such as, for example, lymphokines, interleukin-1 ("IL-1"), interleukin-2 ("IL-2"), interleukin-6 ("IL-6"), granulocyte macrophase colony stimulating factor ("GM-CSF"), granulocyte colony stimulating factor ("G-CSF"), or other growth factors.

Techniques for conjugating such therapeutic moiety to antibodies are well known, see, *e.g.*, Arnon et al., "Monoclonal Antibodies For Immunotargeting Of Drugs In Cancer Therapy", in Monoclonal Antibodies And Cancer Therapy, Reisfeld et al. (eds.), pp. 243-56 (Alan R. Liss, Inc. 1985); Hellstrom et al., "Antibodies For Drug Delivery", in Controlled Drug Delivery (2nd Ed.), Robinson et al. (eds.), pp. 623-53 (Marcel Dekker, Inc. 1987); Thorpe, "Antibody Carriers Of Cytotoxic Agents In Cancer Therapy: A Review", in Monoclonal Antibodies '84: Biological And Clinical Applications, Pinchera et al. (eds.), pp. 475-506 (1985); "Analysis, Results, And Future Prospective Of The Therapeutic Use Of Radiolabeled Antibody In Cancer Therapy", in Monoclonal Antibodies For Cancer Detection And Therapy, Baldwin et al. (eds.), pp. 303-16 (Academic Press 1985), and Thorpe et al., "The Preparation And Cytotoxic Properties Of Antibody-Toxin Conjugates", Immunol. Rev., 62:119-58 (1982).

Alternatively, an antibody can be conjugated to a second antibody to form an antibody heteroconjugate as described by Segal in U.S. Patent No. 4,676,980.

Accordingly, in one aspect, the invention provides substantially purified antibodies or fragment thereof, and non-human antibodies or fragments thereof, which antibodies or fragments specifically bind to a polypeptide comprising an amino acid sequence selected from the group consisting of: the amino acid sequence of any one of SEQ ID Nos. 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 12, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164, or an amino acid sequence encoded by the the cDNA of ATCC® 207192, the cDNA of ATCC® 207223, the cDNA of ATCC® PTA-34, the cDNA of 207189, or the cDNA of ATCC® 207221; a fragment of at least 15 amino acid residues of the amino acid sequence of any one of SEQ ID Nos. 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 12, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164, an amino acid sequence which is at least 95% identical to the amino acid sequence of any one of SEQ ID Nos. 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 12, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164, wherein the percent identity is determined using the

ALIGN program of the GCG software package with a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4; and an amino acid sequence which is encoded by a nucleic acid molecule which hybridizes to the nucleic acid molecule consisting of any one of SEQ ID Nos. 1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 70, 72, 74, 84, 86, 88, 90, 96, 98, 100, 109, 111, 113, 122, 124, 126, 132, 134, 136, 142, 144, 146, 156, 159, 161, or 163, or the cDNA of ATCC® 207192, the cDNA of ATCC® 207223, the cDNA of ATCC® PTA-34, the cDNA of 207189, or the cDNA of ATCC® 207221, or a complement thereof, under conditions of hybridization of 6X SSC at 45°C and washing in 0.2 X SSC, 0.1% SDS at 65°C. In various embodiments, the substantially purified antibodies of the invention, or fragments thereof, can be human, non-human, chimeric and/or humanized antibodies.

In another aspect, the invention provides non-human antibodies or fragments thereof, which antibodies or fragments specifically bind to a polypeptide comprising an amino acid sequence selected from the group consisting of: the amino acid sequence of any one of SEQ ID Nos. 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 12, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164, or an amino acid sequence encoded by the cDNA of ATCC® 207192, the cDNA of ATCC® 207223, the cDNA of ATCC® PTA-34, the cDNA of 207189, or the cDNA of ATCC® 207221; a fragment of at least 15 amino acid residues of the amino acid sequence of any one of SEQ ID Nos. 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 12, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164, an amino acid sequence which is at least 95% identical to the amino acid sequence of any one of SEQ ID Nos. 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 12, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164, wherein the percent identity is determined using the ALIGN program of the GCG software package with a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4; and an amino acid sequence which is encoded by a nucleic acid molecule which hybridizes to the nucleic acid molecule consisting of any one of SEQ ID Nos. 1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 70, 72, 74, 84, 86, 88, 90, 96, 98, 100, 109, 111, 113, 122, 124, 126, 132, 134, 136, 142, 144, 146, 156, 159, 161, or 163, or the cDNA of ATCC® 207192, the cDNA of ATCC® 207223, the cDNA of ATCC® PTA-34, the cDNA of 207189, or the cDNA of ATCC® 207221, or a complement thereof, under conditions of hybridization of 6X SSC at 45°C

and washing in 0.2 X SSC, 0.1% SDS at 65°C. Such non-human antibodies can be goat, mouse, sheep, horse, chicken, rabbit, or rat antibodies. Alternatively, the non-human antibodies of the invention can be chimeric and/or humanized antibodies. In addition, the non-human antibodies of the invention can be polyclonal antibodies or monoclonal antibodies.

In still a further aspect, the invention provides monoclonal antibodies or fragments thereof, which antibodies or fragments specifically bind to a polypeptide comprising an amino acid sequence selected from the group consisting of: the amino acid sequence of any one of SEQ ID Nos. 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 12, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164, or an amino acid sequence encoded by the cDNA of ATCC® 207192, the cDNA of ATCC® 207223, the cDNA of ATCC® PTA-34, the cDNA of 207189, or the cDNA of ATCC® 207221; a fragment of at least 15 amino acid residues of the amino acid sequence of any one of SEQ ID Nos. 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 12, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164, an amino acid sequence which is at least 95% identical to the amino acid sequence of any one of SEQ ID Nos. 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 12, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164, wherein the percent identity is determined using the ALIGN program of the GCG software package with a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4; and an amino acid sequence which is encoded by a nucleic acid molecule which hybridizes to the nucleic acid molecule consisting of any one of SEQ ID Nos. 1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 70, 72, 74, 84, 86, 88, 90, 96, 98, 100, 109, 111, 113, 122, 124, 126, 132, 134, 136, 142, 144, 146, 156, 159, 161, or 163, or the cDNA of ATCC® 207192, the cDNA of ATCC® 207223, the cDNA of ATCC® PTA-34, the cDNA of 207189, or the cDNA of ATCC® 207221, or a complement thereof, under conditions of hybridization of 6X SSC at 45°C and washing in 0.2 X SSC, 0.1% SDS at 65°C. The monoclonal antibodies can be human, humanized, chimeric and/or non-human antibodies.

The substantially purified antibodies or fragments thereof specifically bind to a signal peptide, a secreted sequence, an extracellular domain, a transmembrane or a cytoplasmic domain cytoplasmic membrane of a polypeptide of the invention. In one embodiment, the substantially purified antibodies or fragments thereof, the human or

non-human antibodies or fragments thereof, and/or the monoclonal antibodies or fragments thereof, of the invention specifically bind to a secreted sequence or an extracellular domain of the amino acid sequence of SEQ ID Nos. 5, 11, 17, or 157. Preferably, the secreted sequence or extracellular domain to which the antibody, or fragment thereof, binds comprises from about amino acids 30-362 of SEQ ID NO:5 (SEQ ID NO:42), from amino acids 406-420 of SEQ ID NO:5 (SEQ ID NO:46), from amino acids 19-230 of SEQ ID NO:11 (SEQ ID NO:59), from amino acids 30-362 of SEQ ID NO:17 (SEQ ID NO:65), from amino acids 406-420 of SEQ ID NO:17 (SEQ ID NO:68), and from amino acids 18-229 of SEQ ID NO:157 (SEQ ID NO:171).

Any of the antibodies of the invention can be conjugated to a therapeutic moiety or to a detectable substance. Non-limiting examples of detectable substances that can be conjugated to the antibodies of the invention are an enzyme, a prosthetic group, a fluorescent material, a luminescent material, a bioluminescent material, and a radioactive material.

The invention also provides a kit containing an antibody of the invention conjugated to a detectable substance, and instructions for use. Still another aspect of the invention is a pharmaceutical composition comprising an antibody of the invention and a pharmaceutically acceptable carrier. In one embodiment, the pharmaceutical composition contains an antibody of the invention, a therapeutic moiety, and a pharmaceutically acceptable carrier.

Still another aspect of the invention is a method of making an antibody that specifically recognizes TANGO 204, TANGO 206, TANGO 209, and TANGO A236, the method comprising immunizing a mammal with a polypeptide. The polypeptide used as an immungen comprises an amino acid sequence selected from the group consisting of: the amino acid sequence of any one of SEQ ID Nos. 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 12, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164, or an amino acid sequence encoded by the cDNA of ATCC® 207192, the cDNA of ATCC® 207223, the cDNA of ATCC® PTA-34, the cDNA of 207189, or the cDNA of ATCC® 207221; a fragment of at least 15 amino acid residues of the amino acid sequence of any one of SEQ ID Nos. 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 12, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164, an amino acid sequence which is at least 95% identical to the amino acid sequence of any

one of SEQ ID Nos. 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 12, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164, wherein the percent identity is determined using the ALIGN program of the GCG software package with a PAM120 weight residue table, a gap length penalty of 12, and a gap penalty of 4; and an amino acid sequence which is encoded by a nucleic acid molecule which hybridizes to the nucleic acid molecule consisting of any one of SEQ ID Nos. 1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 70, 72, 74, 84, 86, 88, 90, 96, 98, 100, 109, 111, 113, 122, 124, 126, 132, 134, 136, 142, 144, 146, 156, 159, 161, or 163, or the cDNA of ATCC® 207192, the cDNA of ATCC® 207223, the cDNA of ATCC® PTA-34, the cDNA of 207189, or the cDNA of ATCC® 207221, or a complement thereof, under conditions of hybridization of 6X SSC at 45°C and washing in 0.2 X SSC, 0.1% SDS at 65°C. After immunization, a sample is collected from the mammal that contains an antibody that specifically recognizes a TANGO 204, TANGO 206, TANGO 209, and TANGO A236 polypeptide as exemplified in SEQ ID Nos. 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 12, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164, or portions thereof. Preferably, the polypeptide is recombinantly produced using a non-human host cell. Optionally, the antibodies can be further purified from the sample using techniques well known to those of skill in the art. The method can further comprise producing a monoclonal antibody- producing cell from the cells of the mammal. Optionally, antibodies are collected from the antibody-producing cell.

III. Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors, containing a nucleic acid encoding a polypeptide of the invention (or a portion thereof). As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host

cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors, expression vectors, are capable of directing the expression of genes to which they are operably linked. In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids (vectors). However, the invention is intended to include such other forms of expression vectors, such as viral vectors (e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell. This means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operably linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (e.g., in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (e.g., polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cell and those which direct expression of the nucleotide sequence only in certain host cells (e.g., tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein.

The recombinant expression vectors of the invention can be designed for expression of a polypeptide of the invention in prokaryotic (e.g., *E. coli*) or eukaryotic cells (e.g., insect cells (using baculovirus expression vectors), yeast cells or mammalian cells). Suitable host cells are discussed further in Goeddel, *supra*. Alternatively, the

recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson (1988) *Gene* 67:31-40), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann et al., (1988) *Gene* 69:301-315) and pET 11d (Studier et al., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 60-89). Target gene expression from the pTrc vector relies on host RNA polymerase transcription from a hybrid trp-lac fusion promoter. Target gene expression from the pET 11d vector relies on transcription from a T7 gn10-lac fusion promoter mediated by a coexpressed viral RNA polymerase (T7 gn1). This viral polymerase is supplied by host strains BL21(DE3) or HMS174(DE3) from a resident ? prophage harboring a T7 gn1 gene under the transcriptional control of the lacUV 5 promoter.

One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein (Gottesman, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an

expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (Wada et al. (1992) *Nucleic Acids Res.* 20:2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the expression vector is a yeast expression vector. Examples of vectors for expression in yeast *S. cerivisae* include pYepSec1 (Baldari et al. (1987) *EMBO J.* 6:229-234), pMFa (Kurjan and Herskowitz, (1982) *Cell* 30:933-943), pJRY88 (Schultz et al. (1987) *Gene* 54:113-123), pYES2 (Invitrogen Corporation, San Diego, CA), and pPicZ (Invitrogen Corp, San Diego, CA).

Alternatively, the expression vector is a baculovirus expression vector. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith et al. (1983) *Mol. Cell Biol.* 3:2156-2165) and the pVL series (Lucklow and Summers (1989) *Virology* 170:31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed (1987) *Nature* 329:840) and pMT2PC (Kaufman et al. (1987) *EMBO J.* 6:187-195). When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, Adenovirus 2, cytomegalovirus and Simian Virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see chapters 16 and 17 of Sambrook et al., *supra*.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert et al. (1987) *Genes Dev.* 1:268-277), lymphoid-specific promoters (Calame and Eaton (1988) *Adv. Immunol.* 43:235-275), in particular promoters of T cell receptors (Winoto and Baltimore (1989) *EMBO J.* 8:729-733) and immunoglobulins (Banerji et al. (1983) *Cell* 33:729-740; Queen and Baltimore (1983) *Cell* 33:741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and Ruddle (1989) *Proc. Natl. Acad. Sci. USA* 86:5473-5477), pancreas-specific promoters (Edlund et al. (1985) *Science* 230:912-916),

and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Patent No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, for example the murine hox promoters (Kessel and Gruss (1990) *Science* 249:374-379) and the a-fetoprotein promoter (Campes and Tilghman (1989) *Genes Dev.* 3:537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operably linked to a regulatory sequence in a manner which allows for expression (by transcription of the DNA molecule) of an RNA molecule which is antisense to the mRNA encoding a polypeptide of the invention. Regulatory sequences operably linked to a nucleic acid cloned in the antisense orientation can be chosen which direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen which direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes see Weintraub et al. (*Reviews - Trends in Genetics*, Vol. 1(1) 1986).

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic (e.g., *E. coli*) or eukaryotic cell (e.g., insect cells, yeast or mammalian cells).

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized

techniques for introducing foreign nucleic acid into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, et al. (*supra*), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (e.g., for resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance to drugs, such as G418, hygromycin and methotrexate. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (e.g., cells that have incorporated the selectable marker gene will survive, while the other cells die).

In another embodiment, the expression characteristics of an endogenous (*e.g.*, TANGO 204, TANGO 206, TANGO 209, and A236) nucleic acid within a cell, cell line or microorganism may be modified by inserting a DNA regulatory element heterologous to the endogenous gene of interest into the genome of a cell, stable cell line or cloned microorganism such that the inserted regulatory element is operatively linked with the endogenous gene (*e.g.*, TANGO 204, TANGO 206, TANGO 209, and A236) and controls, modulates or activates the endogenous gene. For example, endogenous TANGO 204, TANGO 206, TANGO 209, and A236 which are normally "transcriptionally silent", *i.e.*, TANGO 204, TANGO 206, TANGO 209, and A236 genes which are normally not expressed, or are expressed only at very low levels in a cell line or microorganism, may be activated by inserting a regulatory element which is capable of promoting the expression of a normally expressed gene product in that cell line or microorganism. Alternatively, transcriptionally silent, endogenous TANGO 204, TANGO 206, TANGO 209, and A236 genes may be activated by insertion of a promiscuous regulatory element that works across cell types.

A heterologous regulatory element may be inserted into a stable cell line or cloned microorganism, such that it is operatively linked with and activates expression of endogenous TANGO 204, TANGO 206, TANGO 209, and A236 genes, using techniques, such as targeted homologous recombination, which are well known to those of skill in the

art, and described *e.g.*, in Chappel, U.S. Patent No. 5,272,071; PCT publication No. WO 91/06667, published May 16, 1991.

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce a polypeptide of the invention. Accordingly, the invention further provides methods for producing a polypeptide of the invention using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding a polypeptide of the invention has been introduced) in a suitable medium such that the polypeptide is produced. In another embodiment, the method further comprises isolating the polypeptide from the medium or the host cell.

The host cells of the invention can also be used to produce nonhuman transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which a sequences encoding a polypeptide of the invention have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous sequences encoding a polypeptide of the invention have been introduced into their genome or homologous recombinant animals in which endogenous encoding a polypeptide of the invention sequences have been altered. Such animals are useful for studying the function and/or activity of the polypeptide and for identifying and/or evaluating modulators of polypeptide activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, etc. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, an "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, e.g., an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing nucleic acid encoding a polypeptide of the invention (or a homologue thereof) into the male pronuclei of a fertilized oocyte, e.g., by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably linked to the transgene to direct expression of the polypeptide of the invention to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, U.S. Patent No. 4,873,191 and in Hogan, *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986) and Wakayama *et al.*, (1999), *Proc. Natl. Acad. Sci. USA*, 96:14984-14989. Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the transgene in its genome and/or expression of mRNA encoding the transgene in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying the transgene can further be bred to other transgenic animals carrying other transgenes.

To create an homologous recombinant animal, a vector is prepared which contains at least a portion of a gene encoding a polypeptide of the invention into which a deletion, addition or substitution has been introduced to thereby alter, e.g., functionally disrupt, the gene. In one embodiment, the vector is designed such that, upon homologous recombination, the endogenous gene is functionally disrupted (i.e., no longer encodes a functional protein; also referred to as a "knock out" vector). Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous gene is mutated or otherwise altered but still encodes functional protein (e.g., the upstream regulatory region can be altered to thereby alter the expression of the endogenous protein). In the homologous recombination vector, the altered portion of the gene is flanked at its 5' and 3' ends by additional nucleic acid of the gene to allow for homologous recombination to occur between the exogenous gene carried by the vector and an endogenous gene in an embryonic stem cell. The additional flanking nucleic acid sequences are of sufficient length for successful homologous recombination with the endogenous gene. Typically,

several kilobases of flanking DNA (both at the 5' and 3' ends) are included in the vector (*see, e.g.,* Thomas and Capecchi (1987) *Cell* 51:503 for a description of homologous recombination vectors). The vector is introduced into an embryonic stem cell line (e.g., by electroporation) and cells in which the introduced gene has homologously recombined with the endogenous gene are selected (*see, e.g.,* Li et al. (1992) *Cell* 69:915). The selected cells are then injected into a blastocyst of an animal (e.g., a mouse) to form aggregation chimeras (*see, e.g.,* Bradley in *Teratocarcinomas and Embryonic Stem Cells: A Practical Approach*, Robertson, ed. (IRL, Oxford, 1987) pp. 113-152). A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley (1991) *Current Opinion in Bio/Technology* 2:823-829 and in PCT Publication NOS. WO 90/11354, WO 91/01140, WO 92/0968, and WO 93/04169.

In another embodiment, transgenic non-human animals can be produced which contain selected systems which allow for regulated expression of the transgene. One example of such a system is the *cre/loxP* recombinase system of bacteriophage P1. For a description of the *cre/loxP* recombinase system, *see, e.g.,* Lakso et al. (1992) *Proc. Natl. Acad. Sci. USA* 89:6232-6236. Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae* (O'Gorman et al. (1991) *Science* 251:1351-1355. If a *cre/loxP* recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the *Cre* recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut et al. (1997) *Nature* 385:810-813 and PCT Publication NOS. WO 97/07668 and WO 97/07669.

IV. Pharmaceutical Compositions

The nucleic acid molecules, polypeptides, and antibodies (also referred to herein as "active compounds") of the invention can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein the language "pharmaceutically acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. The use of such media and agents for pharmaceutically active substances is well known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

The invention includes methods for preparing pharmaceutical compositions for modulating the expression or activity of a polypeptide or nucleic acid of the invention. Such methods comprise formulating a pharmaceutically acceptable carrier with an agent which modulates expression or activity of a polypeptide or nucleic acid of the invention. Such compositions can further include additional active agents. Thus, the invention further includes methods for preparing a pharmaceutical composition by formulating a pharmaceutically acceptable carrier with an agent which modulates expression or activity of a polypeptide or nucleic acid of the invention and one or more addtional active compounds.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, e.g., intravenous, intradermal, subcutaneous, oral (e.g., inhalation), transdermal (topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid; buffers such as acetates, citrates or phosphates and agents for the adjustment of tonicity such as sodium chloride or dextrose. pH can be adjusted with acids or bases, such as hydrochloric acid or

sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersions. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor ELJ (BASF; Parsippany, NJ) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyetheylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as mannitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (e.g., a polypeptide or antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle which contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, the preferred methods of preparation are vacuum drying and freeze-drying which yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral

therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed.

Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from a pressurized container or dispenser which contains a suitable propellant, e.g., a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (e.g., with conventional suppository bases such as cocoa butter and other glycerides) or retention enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides, polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be

obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active compound for the treatment of individuals.

For antibodies, the preferred dosage is 0.1 mg/kg to 100 mg/kg of body weight (generally 10 mg/kg to 20 mg/kg). If the antibody is to act in the brain, a dosage of 50 mg/kg to 100 mg/kg is usually appropriate. Generally, partially human antibodies and fully human antibodies have a longer half-life within the human body than other antibodies. Accordingly, lower dosages and less frequent administration is often possible. Modifications such as lipidation can be used to stabilize antibodies and to enhance uptake and tissue penetration (e.g., into the brain). A method for lipidation of antibodies is described by Cruikshank et al. ((1997) *J. Acquired Immune Deficiency Syndromes and Human Retrovirology* 14:193).

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (U.S. Patent 5,328,470) or by stereotactic injection (*see, e.g.,* Chen et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector can be produced intact from recombinant cells, e.g. retroviral vectors, the

pharmaceutical preparation can include one or more cells which produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

V. Uses and Methods of the Invention

The nucleic acid molecules, proteins, protein homologues, and antibodies described herein can be used in one or more of the following methods: a) screening assays; b) detection assays (e.g., chromosomal mapping, tissue typing, forensic biology); c) predictive medicine (e.g., diagnostic assays, prognostic assays, monitoring clinical trials, and pharmacogenomics); and d) methods of treatment (e.g., therapeutic and prophylactic). For example, the TANGO 221, TANGO 222, TANGO 176, and TANGO 232 polypeptides of the invention can to used to modulate cellular function, survival, morphology, proliferation, and/or differentiation of the cells in which they are expressed. The isolated nucleic acid molecules of the invention can be used to express proteins (e.g., via a recombinant expression vector in a host cell in gene therapy applications), to detect mRNA (e.g., in a biological sample) or a genetic lesion, and to modulate activity of a polypeptide of the invention. In addition, the polypeptides of the invention can be used to screen drugs or compounds which modulate activity or expression of a polypeptide of the invention as well as to treat disorders characterized by insufficient or excessive production of a protein of the invention or production of a form of a protein of the invention which has decreased or aberrant activity compared to the wild type protein. In addition, the antibodies of the invention can be used to detect and isolate a protein of the and modulate activity of a protein of the invention.

This invention further pertains to novel agents identified by the above-described screening assays and uses thereof for treatments as described herein.

A. Screening Assays

The invention provides a method (also referred to herein as a "screening assay") for identifying modulators, i.e., candidate or test compounds or agents (e.g., peptides, peptidomimetics, small molecules or other drugs) which bind to polypeptide of

the invention or have a stimulatory or inhibitory effect on, for example, expression or activity of a polypeptide of the invention.

In one embodiment, the invention provides assays for screening candidate or test compounds which bind to or modulate the activity of the membrane-bound form of a polypeptide of the invention or biologically active portion thereof. The test compounds of the present invention can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the "one-bead one-compound" library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds (Lam (1997) *Anticancer Drug Des.* 12:145).

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt et al. (1993) *Proc. Natl. Acad. Sci. USA* 90:6909; Erb et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:11422; Zuckermann et al. (1994). *J. Med. Chem.* 37:2678; Cho et al. (1993) *Science* 261:1303; Carrell et al. (1994) *Angew. Chem. Int. Ed. Engl.* 33:2059; Carell et al. (1994) *Angew. Chem. Int. Ed. Engl.* 33:2061; and Gallop et al. (1994) *J. Med. Chem.* 37:1233.

Libraries of compounds may be presented in solution (e.g., Houghten (1992) *Bio/Techniques* 13:412-421), or on beads (Lam (1991) *Nature* 354:82-84), chips (Fodor (1993) *Nature* 364:555-556), bacteria (U.S. Patent No. 5,223,409), spores (Patent NOS. 5,571,698; 5,403,484; and 5,223,409), plasmids (Cull et al. (1992) *Proc. Natl. Acad. Sci. USA* 89:1865-1869) or phage (Scott and Smith (1990) *Science* 249:386-390; Devlin (1990) *Science* 249:404-406; Cwirla et al. (1990) *Proc. Natl. Acad. Sci. USA* 87:6378-6382; and Felici (1991) *J. Mol. Biol.* 222:301-310).

In one embodiment, an assay is a cell-based assay in which a cell which expresses a membrane-bound form of a polypeptide of the invention, or a biologically active portion thereof, on the cell surface is contacted with a test compound and the ability of the test compound to bind to the polypeptide determined. The cell, for example, can be a yeast cell or a cell of mammalian origin. Determining the ability of the test compound to bind to the polypeptide can be accomplished, for example, by coupling the test

compound with a radioisotope or enzymatic label such that binding of the test compound to the polypeptide or biologically active portion thereof can be determined by detecting the labeled compound in a complex. For example, test compounds can be labeled with ^{125}I, ^{35}S, ^{14}C, or ^{3}H, either directly or indirectly, and the radioisotope detected by direct counting of radioemmission or by scintillation counting. Alternatively, test compounds can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product. In one embodiment, the assay comprises contacting a cell which expresses a membrane-bound form of a polypeptide of the invention, or a biologically active portion thereof, on the cell surface with a known compound which binds the polypeptide to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the polypeptide, wherein determining the ability of the test compound to interact with the polypeptide comprises determining the ability of the test compound to preferentially bind to the polypeptide or a biologically active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a membrane-bound form of a polypeptide of the invention, or a biologically active portion thereof, on the cell surface with a test compound and determining the ability of the test compound to modulate (e.g., stimulate or inhibit) the activity of the polypeptide or biologically active portion thereof. Determining the ability of the test compound to modulate the activity of the polypeptide or a biologically active portion thereof can be accomplished, for example, by determining the ability of the polypeptide protein to bind to or interact with a target molecule.

Determining the ability of a polypeptide of the invention to bind to or interact with a target molecule can be accomplished by one of the methods described above for determining direct binding. As used herein, a "target molecule" is a molecule with which a selected polypeptide (e.g., a polypeptide of the invention binds or interacts with in nature, for example, a molecule on the surface of a cell which expresses the selected protein, a molecule on the surface of a second cell, a molecule in the extracellular milieu, a molecule associated with the internal surface of a cell membrane or a cytoplasmic molecule. A target molecule can be a polypeptide of the invention or some other

polypeptide or protein. For example, a target molecule can be a component of a signal transduction pathway which facilitates transduction of an extracellular signal (e.g., a signal generated by binding of a compound to a polypeptide of the invention) through the cell membrane and into the cell or a second intercellular protein which has catalytic activity or a protein which facilitates the association of downstream signaling molecules with a polypeptide of the invention. Determining the ability of a polypeptide of the invention to bind to or interact with a target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (e.g., intracellular Ca^{2+}, diacylglycerol, IP3, etc.), detecting catalytic/enzymatic activity of the target on an appropriate substrate, detecting the induction of a reporter gene (e.g., a regulatory element that is responsive to a polypeptide of the invention operably linked to a nucleic acid encoding a detectable marker, e.g. luciferase), or detecting a cellular response, for example, cellular differentiation, or cell proliferation.

In yet another embodiment, an assay of the present invention is a cell-free assay comprising contacting a polypeptide of the invention or biologically active portion thereof with a test compound and determining the ability of the test compound to bind to the polypeptide or biologically active portion thereof. Binding of the test compound to the polypeptide can be determined either directly or indirectly as described above. In one embodiment, the assay includes contacting the polypeptide of the invention or biologically active portion thereof with a known compound which binds the polypeptide to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the polypeptide, wherein determining the ability of the test compound to interact with the polypeptide comprises determining the ability of the test compound to preferentially bind to the polypeptide or biologically active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-free assay comprising contacting a polypeptide of the invention or biologically active portion thereof with a test compound and determining the ability of the test compound to modulate (e.g., stimulate or inhibit) the activity of the polypeptide or biologically active portion thereof. Determining the ability of the test compound to modulate the activity of the polypeptide can be accomplished, for example, by determining the ability of the polypeptide to bind to a

target molecule by one of the methods described above for determining direct binding. In an alternative embodiment, determining the ability of the test compound to modulate the activity of the polypeptide can be accomplished by determining the ability of the polypeptide of the invention to further modulate the target molecule. For example, the catalytic/enzymatic activity of the target molecule on an appropriate substrate can be determined as previously described.

In yet another embodiment, the cell-free assay comprises contacting a polypeptide of the invention or biologically active portion thereof with a known compound which binds the polypeptide to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with the polypeptide, wherein determining the ability of the test compound to interact with the polypeptide comprises determining the ability of the polypeptide to preferentially bind to or modulate the activity of a target molecule.

The cell-free assays of the present invention are amenable to use of both a soluble form or the membrane-bound form of a polypeptide of the invention. In the case of cell-free assays comprising the membrane-bound form of the polypeptide, it may be desirable to utilize a solubilizing agent such that the membrane-bound form of the polypeptide is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside, n-dodecylglucoside, n-octylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton X-100, Triton X-114, Thesit, Isotridecypoly(ethylene glycol ether)n, 3-[(3-cholamidopropyl)dimethylamminio]-1-propane sulfonate (CHAPS), 3-[(3-cholamidopropyl)dimethylamminio]-2-hydroxy-1-propane sulfonate (CHAPSO), or N-dodecyl=N,N-dimethyl-3-ammonio-1-propane sulfonate.

In more than one embodiment of the above assay methods of the present invention, it may be desirable to immobilize either the polypeptide of the invention or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to the polypeptide, or interaction of the polypeptide with a target molecule in the presence and absence of a candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtitre plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be

provided which adds a domain that allows one or both of the proteins to be bound to a matrix. For example, glutathione-S-transferase fusion proteins or glutathione-S-transferase fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical; St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the test compound or the test compound and either the non-adsorbed target protein or A polypeptide of the invention, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads or microtitre plate wells are washed to remove any unbound components and complex formation is measured either directly or indirectly, for example, as described above. Alternatively, the complexes can be dissociated from the matrix, and the level of binding or activity of the polypeptide of the invention can be determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either the polypeptide of the invention or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated polypeptide of the invention or target molecules can be prepared from biotin-NHS (N-hydroxy-succinimide) using techniques well known in the art (e.g., biotinylation kit, Pierce Chemicals; Rockford, IL), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with the polypeptide of the invention or target molecules but which do not interfere with binding of the polypeptide of the invention to its target molecule can be derivatized to the wells of the plate, and unbound target or polypeptide of the invention trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the polypeptide of the invention or target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the polypeptide of the invention or target molecule.

In another embodiment, modulators of expression of a polypeptide of the invention are identified in a method in which a cell is contacted with a candidate compound and the expression of the selected mRNA or protein (i.e., the mRNA or protein corresponding to a polypeptide or nucleic acid of the invention) in the cell is determined. The level of expression of the selected mRNA or protein in the presence of the candidate

compound is compared to the level of expression of the selected mRNA or protein in the absence of the candidate compound. The candidate compound can then be identified as a modulator of expression of the polypeptide of the invention based on this comparison. For example, when expression of the selected mRNA or protein is greater (statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of the selected mRNA or protein expression. Alternatively, when expression of the selected mRNA or protein is less (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of the selected mRNA or protein expression. The level of the selected mRNA or protein expression in the cells can be determined by methods described herein.

In yet another aspect of the invention, a polypeptide of the inventions can be used as "bait proteins" in a two-hybrid assay or three hybrid assay (*see, e.g.*, U.S. Patent No. 5,283,317; Zervos et al. (1993) *Cell* 72:223-232; Madura et al. (1993) *J. Biol. Chem.* 268:12046-12054; Bartel et al. (1993) *Bio/Techniques* 14:920-924; Iwabuchi et al. (1993) *Oncogene* 8:1693-1696; and PCT Publication No. WO 94/10300), to identify other proteins, which bind to or interact with the polypeptide of the invention and modulate activity of the polypeptide of the invention. Such binding proteins are also likely to be involved in the propagation of signals by the polypeptide of the inventions as, for example, upstream or downstream elements of a signaling pathway involving the polypeptide of the invention.

This invention further pertains to novel agents identified by the above-described screening assays and uses thereof for treatments as described herein.

B. Detection Assays

Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. For example, these sequences can be used to: (i) map their respective genes on a chromosome and, thus, locate gene regions associated with genetic disease; (ii) identify an individual from a minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample. These applications are described in the subsections below.

1. Chromosome Mapping

Once the sequence (or a portion of the sequence) of a gene has been isolated, this sequence can be used to map the location of the gene on a chromosome. Accordingly, nucleic acid molecules described herein or fragments thereof, can be used to map the location of the corresponding genes on a chromosome. The mapping of the sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

Briefly, genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the sequence of a gene of the invention. Computer analysis of the sequence of a gene of the invention can be used to rapidly select primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the gene sequences will yield an amplified fragment. For a review of this technique, see D'Eustachio et al. ((1983) *Science* 220:919-924).

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be assigned per day using a single thermal cycler. Using the nucleic acid sequences of the invention to design oligonucleotide primers, sublocalization can be achieved with panels of fragments from specific chromosomes. Other mapping strategies which can similarly be used to map a gene to its chromosome include *in situ* hybridization (described in Fan et al. (1990) *Proc. Natl. Acad. Sci. USA* 87:6223-27), pre-screening with labeled flow-sorted chromosomes (CITE), and pre-selection by hybridization to chromosome specific cDNA libraries. Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. For a review of this technique, see Verma et al., (Human Chromosomes: A Manual of Basic Techniques (Pergamon Press, New York, 1988)).

Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding

sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations during chromosomal mapping.

Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. (Such data are found, for example, in V. McKusick, Mendelian Inheritance in Man, available on-line through Johns Hopkins University Welch Medical Library). The relationship between genes and disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, e.g., Egeland et al. (1987) *Nature* 325:783-787.

Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with a gene of the invention can be determined. If a mutation is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

Furthermore, the nucleic acid sequences disclosed herein can be used to perform searches against "mapping databases", *e.g.*, BLAST-type search, such that the chromosome position of the gene is identified by sequence homology or identity with known sequence fragments which have been mapped to chromosomes.

In the instant case, the mouse TANGO 204 gene was mapped to mouse using the Genebridge 4 Radiation hybrid mapping panel with gacaagctgcattcaaagcttcc (SEQ ID NO:82) as the forward primer and ctggagcacatggtagtgattc (SEQ ID NO:83) as the reverse primer. The mouse TANGO 204 gene maps to chromosome 1. Flanking markers for this region are D1Mit430 and D1Mit119. Mapping by synteny reveals that human TANGO 204 maps to human chromosome 8q. The CCAL1 (chondrocalcinosis 1) locus also maps to this region of the human chromosome. The OPRK (opiate receptor) gene also maps to this region of the human chromosome. The tb (tumbler), fz (fuzzy) loci also map to this region of the mouse chromosome. The tb (tumbler), fz (fuzzy) genes also map to this region of the mouse chromosome.

In the instant case, the human gene for TANGO 209 was mapped on radiation hybrid panels to the long arm of chromosome 6, in the region q26-27. Flanking markers for this region are ATA22G07 and WI-9405. The MLLT4(myeloid/lymphoid or mixed lineage leukemia) locus also maps to this region of the human chromosome. The PLG (plasminogen), VIP (vasoactive intestinal peptide), LPA (apolipoprotein Lp), MLLT4 (myeloid/lymphoid or mixed lineage leukemia), and THBS2 (thrombospondin 2) genes also map to this region of the human chromosome. This region is syntenic to mouse chromosome 17. The qk (quaking), T (brachyury), and het (head tilt) loci also map to this region of the mouse chromosome. The plg (plasminogen), qk (quaking), and het (head tilt) genes also map to this region of the mouse chromosome.

A polypeptide and fragments and sequences thereof and antibodies specific thereto can be used to map the location of the gene encoding the polypeptide on a chromosome. This mapping can be carried out by specifically detecting the presence of the polypeptide in members of a panel of somatic cell hybrids between cells of a first species of animal from which the protein originates and cells from a second species of animal and then determining which somatic cell hybrid(s) expresses the polypeptide and noting the chromosome(s) from the first species of animal that it contains. For examples of this technique, see Pajunen *et al.* (1988*) Cytogenet. Cell Genet. 47*:37-41 and Van Keuren *et al.* (1986*) Hum. Genet. 74*:34-40. Alternatively, the presence of the polypeptide in the somatic cell hybrids can be determined by assaying an activity or property of the polypeptide, for example, enzymatic activity, as described in Bordelon-Riser *et al.* (1979*) Somatic Cell Genetics 5*:597-613 and Owerbach *et al.* (1978*) Proc. Natl. Acad. Sci. USA 75*:5640-5644.

2. Tissue Typing

The nucleic acid sequences of the present invention can also be used to identify individuals from minute biological samples. The United States military, for example, is considering the use of restriction fragment length polymorphism (RFLP) for identification of its personnel. In this technique, an individual's genomic DNA is digested with one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. This method does not suffer from the current limitations of "Dog Tags" which can be lost, switched, or stolen, making positive identification difficult. The

sequences of the present invention are useful as additional DNA markers for RFLP (described in U.S. Patent 5,272,057).

Furthermore, the sequences of the present invention can be used to provide an alternative technique which determines the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the nucleic acid sequences described herein can be used to prepare two PCR primers from the 5' and 3' ends of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the present invention can be used to obtain such identification sequences from individuals and from tissue. The nucleic acid sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences of SEQ ID NO:1, 4, 7, and 10 can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers which each yield a noncoding amplified sequence of 100 bases. If predicted coding sequences, such as those in SEQ ID NO:3, 6, 9, and 12, are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

If a panel of reagents from the nucleic acid sequences described herein is used to generate a unique identification database for an individual, those same reagents can later be used to identify tissue from that individual. Using the unique identification database, positive identification of the individual, living or dead, can be made from extremely small tissue samples.

3. Use of Partial Gene Sequences in Forensic Biology

DNA-based identification techniques can also be used in forensic biology. Forensic biology is a scientific field employing genetic typing of biological evidence found at a crime scene as a means for positively identifying, for example, a perpetrator of a crime. To make such an identification, PCR technology can be used to amplify DNA sequences taken from very small biological samples such as tissues, e.g., hair or skin, or body fluids, e.g., blood, saliva, or semen found at a crime scene. The amplified sequence can then be compared to a standard, thereby allowing identification of the origin of the biological sample.

The sequences of the present invention can be used to provide polynucleotide reagents, e.g., PCR primers, targeted to specific loci in the human genome, which can enhance the reliability of DNA-based forensic identifications by, for example, providing another "identification marker" (i.e. another DNA sequence that is unique to a particular individual). As mentioned above, actual base sequence information can be used for identification as an accurate alternative to patterns formed by restriction enzyme generated fragments. Sequences targeted to noncoding regions are particularly appropriate for this use as greater numbers of polymorphisms occur in the noncoding regions, making it easier to differentiate individuals using this technique. Examples of polynucleotide reagents include the nucleic acid sequences of the invention or portions thereof, e.g., fragments derived from noncoding regions having a length of at least 20 or 30 bases.

The nucleic acid sequences described herein can further be used to provide polynucleotide reagents, e.g., labeled or labelable probes which can be used in, for example, an *in situ* hybridization technique, to identify a specific tissue, e.g., brain tissue. This can be very useful in cases where a forensic pathologist is presented with a tissue of unknown origin. Panels of such probes can be used to identify tissue by species and/or by organ type.

C. Predictive Medicine

The present invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the present invention relates to diagnostic assays for determining TANGO 204, TANGO 206, TANGO 209, and A236 protein and/or nucleic acid expression as well as

TANGO 204, TANGO 206, TANGO 209, and A236 activity, in the context of a biological sample (e.g., blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or disorder, or is at risk of developing a disorder, associated with aberrant or unwanted TANGO 204, TANGO 206, TANGO 209, and A236 expression or activity. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with TANGO 204, TANGO 206, TANGO 209, and A236 protein, nucleic acid expression or activity. For example, mutations in a TANGO 204, TANGO 206, TANGO 209, and A236 gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with TANGO 204, TANGO 206, TANGO 209, and A236 protein, nucleic acid expression or activity.

As an alternative to making determinations based on the absolute expression level of selected genes, determinations may be based on the normalized expression levels of these genes. Expression levels are normalized by correcting the absolute expression level of a TANGO 204, TANGO 206, TANGO 209, and A236 gene by comparing its expression to the expression of a gene that is not a TANGO 204, TANGO 206, TANGO 209, and A236 gene, e.g., a housekeeping gene that is constitutively expressed. Suitable genes for normalization include housekeeping genes such as the actin gene. This normalization allows the comparison of the expression level in one sample, e.g., a patient sample, to another sample, e.g., a non-[disease] sample, or between samples from different sources.

Alternatively, the expression level can be provided as a relative expression level. To determine a relative expression level of a gene, the level of expression of the gene is determined for 10 or more samples of different cell isolates, preferably 50 or more samples, prior to the determination of the expression level for the sample in question. The cell isolates are selected depending upon the tissues in which the gene of interest is expressed. For example, for TANGO 204 family members, expression was observed in the eye, lung, stomach, intestine and developing bone and cartilage structures such as the ear, nose, and spinal column. For TANGO 206 family members, expression was observed in the heart, skeletal muscle, kidney, brain, and placenta, central nervous system, eye, large ganglion of the head, liver and skin. For TANGO 209 family members, expression

was observed in the heart, skeletal muscle, pancreas, placenta, lung, kidney, brain (hippocampus, dentate gyrus, and frontal cortex), thymus (multifocal expression), kidney (medulla and capsule), adrenal gland (capsule), spleen (non-follicular), bladder, small intestine and colon (smooth muscle, not villi), liver. For A236 family members, expression was observed in the brain, placenta, uterus, ovary, mandible, and the intestinal tract. The mean expression level of each of the genes assayed in the larger number of samples is determined and this is used as a baseline expression level for the gene(s) in question. The expression level of the gene determined for the test sample (absolute level of expression) is then divided by the mean expression value obtained for that gene. This provides a relative expression level and aids in identifying extreme cases of TANGO 204, TANGO 206, TANGO 209, TANGO A236-mediated disease.

For example, by way of illustration only, for TANGO 204 family members, diseases which may be studied include, without limitation, those associated with tissues of the eye, lung, stomach, intestine and developing bone and cartilage structures such as the ear, nose, and spinal column. For TANGO 206 family members, diseases which may be studied include, without limitation, those associated with tissues of the heart, skeletal muscle, kidney, brain, and placenta, central nervous system, eye, large ganglion of the head, liver and skin. For TANGO 209 family members, diseases which may be studied include, without limitation, those associated with tissues of the heart, skeletal muscle, pancreas, placenta, lung, kidney, brain (hippocampus, dentate gyrus, and frontal cortex), thymus (multifocal expression), kidney (medulla and capsule), adrenal gland (capsule), spleen (non-follicular), bladder, small intestine and colon (smooth muscle, not villi), liver. For A236 family members, diseases which may be studied include, without limitation, those associated with tissues of the brain, placenta, uterus, ovary, mandible, and the intestinal tract.

Preferably, the samples used in the baseline determination will be from TANGO 204, TANGO 206, TANGO 209, and A236-mediated diseased or from non-diseased cells of tissue. The choice of the cell source is dependent on the use of the relative expression level. Using expression found in normal tissues as a mean expression score aids in validating whether the TANGO 204, TANGO 206, TANGO 209, and A236 gene assayed is cell-type specific for the tissues in which expression is observed versus the expression found in normal cells. Such a use is particularly important in identifying

whether a TANGO 204, TANGO 206, TANGO 209, and A236 gene can serve as a target gene. In addition, as more data is accumulated, the mean expression value can be revised, providing improved relative expression values based on accumulated data. Expression data from alveoli, osteoblast, bone marrow cells provides a means for grading the severity of the TANGO 204, TANGO 206, TANGO 209, and A236-mediated disease state.

Another aspect of the invention pertains to monitoring the influence of agents (e.g., drugs, compounds) on the expression or activity of TANGO 204, TANGO 206, TANGO 209, and A236 in clinical trials.

These and other agents are described in further detail in the following sections.

1. Diagnostic Assays

An exemplary method for detecting the presence or absence of a polypeptide or nucleic acid of the invention in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting a polypeptide or nucleic acid (e.g., mRNA, genomic DNA) of the invention such that the presence of a polypeptide or nucleic acid of the invention is detected in the biological sample. A preferred agent for detecting mRNA or genomic DNA encoding a polypeptide of the invention is a labeled nucleic acid probe capable of hybridizing to mRNA or genomic DNA encoding a polypeptide of the invention. The nucleic acid probe can be, for example, a full-length cDNA, such as the nucleic acid of SEQ ID Nos. 1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 70, 72, 74, 84, 86, 88, 90, 96, 98, 100, 109, 111, 113, 122, 124, 126, 132, 134, 136, 142, 144, 146, 156, 159, 161, or 163, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to a mRNA or genomic DNA encoding a polypeptide of the invention. Other suitable probes for use in the diagnostic assays of the invention are described herein.

A preferred agent for detecting a polypeptide of the invention is an antibody capable of binding to a polypeptide of the invention, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (e.g., Fab or $F(ab')_2$) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (i.e., physically linking) a detectable substance to the

probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of a polypeptide of the invention include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. *In vitro* techniques for detection of genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of a polypeptide of the invention include introducing into a subject a labeled antibody directed against the polypeptide. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject.

In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting a polypeptide of the invention or mRNA or genomic DNA encoding a polypeptide of the invention, such that the presence of the polypeptide or mRNA or genomic DNA encoding the polypeptide is detected in the biological sample, and comparing the presence of the polypeptide or mRNA or genomic DNA encoding the polypeptide in the control sample with the presence of the polypeptide or mRNA or genomic DNA encoding the polypeptide in the test sample.

The invention also encompasses kits for detecting the presence of a polypeptide or nucleic acid of the invention in a biological sample (a test sample). Such kits can be used to determine if a subject is suffering from or is at increased risk of

developing a disorder associated with aberrant expression of a polypeptide of the invention (e.g., a proliferative disorder, e.g., psoriasis or cancer). For example, the kit can comprise a labeled compound or agent capable of detecting the polypeptide or mRNA encoding the polypeptide in a biological sample and means for determining the amount of the polypeptide or mRNA in the sample (e.g., an antibody which binds the polypeptide or an oligonucleotide probe which binds to DNA or mRNA encoding the polypeptide). Kits can also include instructions for observing that the tested subject is suffering from or is at risk of developing a disorder associated with aberrant expression of the polypeptide if the amount of the polypeptide or mRNA encoding the polypeptide is above or below a normal level.

For antibody-based kits, the kit can comprise, for example: (1) a first antibody (e.g., attached to a solid support) which binds to a polypeptide of the invention; and, optionally, (2) a second, different antibody which binds to either the polypeptide or the first antibody and is conjugated to a detectable agent.

For oligonucleotide-based kits, the kit can comprise, for example: (1) an oligonucleotide, e.g., a detectably labeled oligonucleotide, which hybridizes to a nucleic acid sequence encoding a polypeptide of the invention or (2) a pair of primers useful for amplifying a nucleic acid molecule encoding a polypeptide of the invention. The kit can also comprise, e.g., a buffering agent, a preservative, or a protein stabilizing agent. The kit can also comprise components necessary for detecting the detectable agent (e.g., an enzyme or a substrate). The kit can also contain a control sample or a series of control samples which can be assayed and compared to the test sample contained. Each component of the kit is usually enclosed within an individual container and all of the various containers are within a single package along with instructions for observing whether the tested subject is suffering from or is at risk of developing a disorder associated with aberrant expression of the polypeptide.

2. Prognostic Assays

The methods described herein can furthermore be utilized as diagnostic or prognostic assays to identify subjects having or at risk of developing a disease or disorder associated with aberrant expression or activity of a polypeptide of the invention. For example, the assays described herein, such as the preceding diagnostic assays or the

following assays, can be utilized to identify a subject having or at risk of developing a disorder associated with aberrant expression or activity of a polypeptide of the invention. Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing such a disease or disorder. Thus, the present invention provides a method in which a test sample is obtained from a subject and a polypeptide or nucleic acid (e.g., mRNA, genomic DNA) of the invention is detected, wherein the presence of the polypeptide or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant expression or activity of the polypeptide. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (e.g., serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (e.g., an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant expression or activity of a polypeptide of the invention. For example, such methods can be used to determine whether a subject can be effectively treated with a specific agent or class of agents (e.g., agents of a type which decrease activity of the polypeptide). Thus, the present invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant expression or activity of a polypeptide of the invention in which a test sample is obtained and the polypeptide or nucleic acid encoding the polypeptide is detected (e.g., wherein the presence of the polypeptide or nucleic acid is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant expression or activity of the polypeptide).

The methods of the invention can also be used to detect genetic lesions or mutations in a gene of the invention, thereby determining if a subject with the lesioned gene is at risk for a disorder characterized aberrant expression or activity of a polypeptide of the invention. In one embodiment, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic lesion or mutation characterized by at least one of an alteration affecting the integrity of a gene encoding the polypeptide of the invention, or the mis-expression of the gene encoding the polypeptide of the invention. For example, such genetic lesions or mutations can be detected by ascertaining the

existence of at least one of: 1) a deletion of one or more nucleotides from the gene; 2) an addition of one or more nucleotides to the gene; 3) a substitution of one or more nucleotides of the gene; 4) a chromosomal rearrangement of the gene; 5) an alteration in the level of a messenger RNA transcript of the gene; 6) an aberrant modification of the gene, such as of the methylation pattern of the genomic DNA; 7) the presence of a non-wild type splicing pattern of a messenger RNA transcript of the gene; 8) a non-wild type level of a the protein encoded by the gene; 9) an allelic loss of the gene; and 10) an inappropriate post-translational modification of the protein encoded by the gene. As described herein, there are a large number of assay techniques known in the art which can be used for detecting lesions in a gene.

In certain embodiments, detection of the lesion involves the use of a probe/primer in a polymerase chain reaction (PCR) (*see, e.g.*, U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (*see, e.g.*, Landegran et al. (1988) *Science* 241:1077-1080; and Nakazawa et al. (1994) *Proc. Natl. Acad. Sci. USA* 91:360-364), the latter of which can be particularly useful for detecting point mutations in a gene (*see, e.g.*, Abravaya et al. (1995) *Nucleic Acids Res.* 23:675-682). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to the selected gene under conditions such that hybridization and amplification of the gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (Guatelli et al. (1990) *Proc. Natl. Acad. Sci. USA* 87:1874-1878), transcriptional amplification system (Kwoh, et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:1173-1177), Q-Beta Replicase (Lizardi et al. (1988) *Bio/Technology* 6:1197), or any other nucleic acid amplification method, followed by the detection of the amplified molecules using techniques well known to those of skill in the art. These detection schemes are especially

useful for the detection of nucleic acid molecules if such molecules are present in very low numbers.

In an alternative embodiment, mutations in a selected gene from a sample cell can be identified by alterations in restriction enzyme cleavage patterns. For example, sample and control DNA is isolated, amplified (optionally), digested with one or more restriction endonucleases, and fragment length sizes are determined by gel electrophoresis and compared. Differences in fragment length sizes between sample and control DNA indicates mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (*see, e.g.*, U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations can be identified by hybridizing a sample and control nucleic acids, e.g., DNA or RNA, to high density arrays containing hundreds or thousands of oligonucleotides probes (Cronin et al. (1996) *Human Mutation* 7:244-255; Kozal et al. (1996) *Nature Medicine* 2:753-759). For example, genetic mutations can be identified in two-dimensional arrays containing light-generated DNA probes as described in Cronin et al., *supra*. Briefly, a first hybridization array of probes can be used to scan through long stretches of DNA in a sample and control to identify base changes between the sequences by making linear arrays of sequential overlapping probes. This step allows the identification of point mutations. This step is followed by a second hybridization array that allows the characterization of specific mutations by using smaller, specialized probe arrays complementary to all variants or mutations detected. Each mutation array is composed of parallel probe sets, one complementary to the wild-type gene and the other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art can be used to directly sequence the selected gene and detect mutations by comparing the sequence of the sample nucleic acids with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxim and Gilbert ((1977) *Proc. Natl. Acad. Sci. USA* 74:560) or Sanger ((1977) *Proc. Natl. Acad. Sci. USA* 74:5463). It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays ((1995) *Bio/Techniques* 19:448), including sequencing by mass spectrometry (*see,*

e.g., PCT Publication No. WO 94/16101; Cohen et al. (1996) *Adv. Chromatogr.* 36:127-162; and Griffin et al. (1993) *Appl. Biochem. Biotechnol.* 38:147-159).

Other methods for detecting mutations in a selected gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes (Myers et al. (1985) *Science* 230:1242). In general, the technique of "mismatch cleavage" entails providing heteroduplexes formed by hybridizing (labeled) RNA or DNA containing the wild-type sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent which cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. RNA/DNA duplexes can be treated with RNase to digest mismatched regions, and DNA/DNA hybrids can be treated with S1 nuclease to digest mismatched regions.

In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. *See, e.g.*, Cotton et al. (1988) *Proc. Natl. Acad. Sci. USA* 85:4397; Saleeba et al. (1992) *Methods Enzymol.* 217:286-295. In a one embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches (Hsu et al. (1994) *Carcinogenesis* 15:1657-1662). According to an exemplary embodiment, a probe based on a selected sequence, e.g., a wild-type sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. *See, e.g.*, U.S. Patent No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in genes. For example, single strand conformation polymorphism

(SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids (Orita et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:2766; *see also* Cotton (1993) *Mutat. Res.* 285:125-144; Hayashi (1992) *Genet. Anal. Tech. Appl.* 9:73-79). Single-stranded DNA fragments of sample and control nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, and the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In one embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility (Keen et al. (1991) *Trends Genet.* 7:5).

In yet another embodiment, the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE) (Myers et al. (1985) *Nature* 313:495). When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a `GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA (Rosenbaum and Reissner (1987) *Biophys. Chem.* 265:12753).

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions which permit hybridization only if a perfect match is found (Saiki et al. (1986) *Nature* 324:163); Saiki et al. (1989) *Proc. Natl. Acad. Sci. USA* 86:6230). Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology which depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of

interest in the center of the molecule (so that amplification depends on differential hybridization) (Gibbs et al. (1989) *Nucleic Acids Res.* 17:2437-2448) or at the extreme 3' end of one primer where, under appropriate conditions, mismatch can prevent or reduce polymerase extension (Prossner (1993) *Tibtech* 11:238). In addition, it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection (Gasparini et al. (1992) *Mol. Cell Probes* 6:1). It is anticipated that in certain embodiments amplification may also be performed using Taq ligase for amplification (Barany (1991) *Proc. Natl. Acad. Sci. USA* 88:189). In such cases, ligation will occur only if there is a perfect match at the 3' end of the 5' sequence making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, e.g., in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a gene encoding a polypeptide of the invention. Furthermore, any cell type or tissue, preferably peripheral blood leukocytes, in which the polypeptide of the invention is expressed may be utilized in the prognostic assays described herein.

3. Pharmacogenomics

Agents, or modulators which have a stimulatory or inhibitory effect on activity or expression of a polypeptide of the invention as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) disorders associated with aberrant activity of the polypeptide. In conjunction with such treatment, the pharmacogenomics (i.e., the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) of the individual may be considered. Differences in metabolism of therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, the pharmacogenomics of the individual permits the selection of effective agents (e.g., drugs) for prophylactic or therapeutic treatments based on a consideration of the individual's genotype. Such pharmacogenomics can further be used to determine appropriate dosages and therapeutic

regimens. Accordingly, the activity of a polypeptide of the invention, expression of a nucleic acid of the invention, or mutation content of a gene of the invention in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual.

Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. *See, e.g.*, Linder (1997) *Clin. Chem.* 43(2):254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body are referred to as "altered drug action." Genetic conditions transmitted as single factors altering the way the body acts on drugs are referred to as "altered drug metabolism". These pharmacogenetic conditions can occur either as rare defects or as polymorphisms. For example, glucose-6-phosphate dehydrogenase deficiency (G6PD) is a common inherited enzymopathy in which the main clinical complication is haemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (e.g., N-acetyltransferase 2 (NAT 2) and cytochrome P450 enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, a PM will show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite morphine. The other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Thus, the activity of a polypeptide of the invention, expression of a nucleic acid encoding the polypeptide, or mutation content of a gene encoding the polypeptide in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual. In addition, pharmacogenetic studies can be used to apply genotyping of polymorphic alleles encoding drug-metabolizing enzymes to the identification of an individual's drug responsiveness phenotype. This knowledge, when applied to dosing or drug selection, can avoid adverse reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a modulator of activity or expression of the polypeptide, such as a modulator identified by one of the exemplary screening assays described herein.

4. Monitoring of Effects During Clinical Trials

Monitoring the influence of agents (e.g., drugs, compounds) on the expression or activity of a polypeptide of the invention (e.g., the ability to modulate aberrant cell proliferation chemotaxis, and/or differentiation) can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent, as determined by a screening assay as described herein, to increase gene expression, protein levels or protein activity, can be monitored in clinical trials of subjects exhibiting decreased gene expression, protein levels, or protein activity. Alternatively, the effectiveness of an agent, as determined by a screening assay, to decrease gene expression, protein levels or protein activity, can be monitored in clinical trials of subjects exhibiting increased gene expression, protein levels, or protein activity. In such clinical trials, expression or activity of a polypeptide of the invention and preferably, that of other polypeptide that have been implicated in for example, a cellular proliferation disorder, can be used as a marker of the immune responsiveness of a particular cell.

For example, and not by way of limitation, genes, including those of the invention, that are modulated in cells by treatment with an agent (e.g., compound, drug or small molecule) which modulates activity or expression of a polypeptide of the invention (e.g., as identified in a screening assay described herein) can be identified. Thus, to study the effect of agents on cellular proliferation disorders, for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of a gene of the invention and other genes implicated in the disorder. The levels of gene expression

(i.e., a gene expression pattern) can be quantified by Northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of a gene of the invention or other genes. In this way, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during, treatment of the individual with the agent.

In a preferred embodiment, the present invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (e.g., an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) comprising the steps of (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of the polypeptide or nucleic acid of the invention in the preadministration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level the of the polypeptide or nucleic acid of the invention in the post-administration samples; (v) comparing the level of the polypeptide or nucleic acid of the invention in the pre-administration sample with the level of the polypeptide or nucleic acid of the invention in the post-administration sample or samples; and (vi) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of the polypeptide to higher levels than detected, i.e., to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease expression or activity of the polypeptide to lower levels than detected, i.e., to decrease the effectiveness of the agent.

C. Methods of Treatment

The present invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant expression or activity of a polypeptide of the invention, e.g., cardiac infection (e.g., myocarditis or dilated cardiomyopathy), central nervous system infection (e.g., non-specific ferbile illness or meningoencephalitis), pancreatic infection (e.g., acute pancreatitis), respiratory infection (pneumonia), gastrointestinal infection, type I diabetes,

cancer, familia hypercholesterolemia, treat hemophilia B, Marfan syndrome, protein S deficiency, allergy, inflammation, and gastroduodenal ulcer. Moreover, the polypeptides of the invention can be used to modulate cellular function, survival, morphology, proliferation and/or differentiation.

1. Prophylactic Methods

In one aspect, the invention provides a method for preventing in a subject, a disease or condition associated with an aberrant expression or activity of a polypeptide of the invention, by administering to the subject an agent which modulates expression or at least one activity of the polypeptide. Subjects at risk for a disease which is caused or contributed to by aberrant expression or activity of a polypeptide of the invention can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the manifestation of symptoms characteristic of the aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending on the type of aberrancy, for example, an agonist or antagonist agent can be used for treating the subject.

2. Therapeutic Methods

Another aspect of the invention pertains to methods of modulating expression or activity of a polypeptide of the invention for therapeutic purposes. The modulatory method of the invention involves contacting a cell with an agent that modulates one or more of the activities of the polypeptide. An agent that modulates activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring cognate ligand of the polypeptide, a peptide, a peptidomimetic, or other small molecule. In one embodiment, the agent stimulates one or more of the biological activities of the polypeptide. Examples of such stimulatory agents include the active polypeptide of the invention and a nucleic acid molecule encoding the polypeptide of the invention that has been introduced into the cell. In another embodiment, the agent inhibits one or more of the biological activities of the polypeptide of the invention. Examples of such inhibitory agents include antisense nucleic acid molecules and antibodies. These modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g, by administering the agent to a subject). As such, the present

invention provides methods of treating an individual afflicted with a disease or disorder characterized by aberrant expression or activity of a polypeptide of the invention. In one embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination of agents that modulates (e.g., upregulates or downregulates) expression or activity. In another embodiment, the method involves administering a polypeptide of the invention or a nucleic acid molecule of the invention as therapy to compensate for reduced or aberrant expression or activity of the polypeptide.

Stimulation of activity is desirable in situations in which activity or expression is abnormally low or downregulated and/or in which increased activity is likely to have a beneficial effect. Conversely, inhibition of activity is desirable in situations in which activity or expression is abnormally high or upregulated and/or in which decreased activity is likely to have a beneficial effect.

All publications, patents and patent applications mentioned in this specification are herein incorporated by reference into the specification to the same extent as if each individual publication, patent or patent application was specifically and individually indicated to be incorporated herein by reference.

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

International Application No: PCT/ /

MICROORGANISMS

Optional Sheet in connection with the microorganism referred to on page 67-68, lines 1-30; 1-18 of the description

A. IDENTIFICATION OF DEPOSIT •

Further deposits are identified on an additional sheet •

Name of depositary institution •

American Type Culture Collection

Address of depositary institution (including postal code and country) •

10801 University Blvd.
Manassas, VA 20110-2209
US

Date of deposit • April 2, 1999 Accession Number • 207189

B. ADDITIONAL INDICATIONS • (leave blank if not applicable). This information is continued on a separate attached sheet

C. DESIGNATED STATES FOR WHICH INDICATIONS ARE MADE • (if the indications are not all designated States)

D. SEPARATE FURNISHING OF INDICATIONS • (leave blank if not applicable)

The indications listed below will be submitted to the International Bureau later • (Specify the general nature of the indications e.g., "Accession Number of Deposit")

E. ☑ This sheet was received with the International application when filed (to be checked by the receiving Office)

SIGFRED HOSTAD
INTERNATIONAL DIVISION
703 305 3680

☐ The date of receipt (from the applicant) by the International Bureau

was

(Authorized Officer)

Form PCT/RO/134 (January 1981)

International Application No: PCT/ /

Form PCT/RO/134 (cont.)

American Type Culture Collection

10801 University Blvd.,
Manassas, VA 20110-2209
US

Accession No.	Date of Deposit
207192	April 2, 1999
207221	April 21, 1999
207223	April 21, 1999
PTA-34	May 7, 1999

WHAT IS CLAIMED IS:

1. An isolated nucleic acid molecule selected from the group consisting of:

a) a nucleic acid molecule comprising a nucleotide sequence which is at least 55% identical to the nucleotide sequence of SEQ ID Nos. 1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 70, 72, 74, 84, 86, 88, 90, 96, 98, 100, 109, 111, 113, 122, 124, 126, 132, 134, 136, 142, 144, 146, 156, 159, 161, or 163, the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207192, the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207223, the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-34, the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207189,the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207221, or a complement thereof;

b) a nucleic acid molecule comprising a fragment of at least 300 nucleotides of the nucleotide sequence of SEQ ID Nos. 1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 70, 72, 74, 84, 86, 88, 90, 96, 98, 100, 109, 111, 113, 122, 124, 126, 132, 134, 136, 142, 144, 146, 156, 159, 161, or 163, the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207192, the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207223, the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-34, the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207189,the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207221, or a complement thereof;

c) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO: 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 112, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207192, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207223, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-34, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207189, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207221;

d) a nucleic acid molecule which encodes a fragment of a polypeptide comprising the amino acid sequence of SEQ ID NO: 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 112, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207192, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207223, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-34, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207189, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207221, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID NO: 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 112, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207192, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207223, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-34, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207189, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207221; and

e) a nucleic acid molecule which encodes a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID NO: 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 112, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207192, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207223, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-34, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207189, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207221, wherein the nucleic

acid molecule hybridizes to a nucleic acid molecule comprising SEQ ID Nos. 1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 70, 72, 74, 84, 86, 88, 90, 96, 98, 100, 109, 111, 113, 122, 124, 126, 132, 134, 136, 142, 144, 146, 156, 159, 161, or 163, or a complement thereof, under stringent conditions.

2. The isolated nucleic acid molecule of claim 1, which is selected from the group consisting of:

a) a nucleic acid comprising the nucleotide sequence of SEQ ID Nos. 1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 70, 72, 74, 84, 86, 88, 90, 96, 98, 100, 109, 111, 113, 122, 124, 126, 132, 134, 136, 142, 144, 146, 156, 159, 161, or 163, the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207192, the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207223, the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-34, the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207189,the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207221, or a complement thereof; and

b) a nucleic acid molecule which encodes a polypeptide comprising the amino acid sequence of SEQ ID NO: 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 112, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207192, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207223, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-34, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207189, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207221.

3. The nucleic acid molecule of claim 1 further comprising vector nucleic acid sequences.

4. The nucleic acid molecule of claim 1 further comprising nucleic acid sequences encoding a heterologous polypeptide.

5. A host cell which contains the nucleic acid molecule of claim 1.

6. The host cell of claim 5 which is a mammalian host cell.

7. A non-human mammalian host cell containing the nucleic acid molecule of claim 1.

8. An isolated polypeptide selected from the group consisting of:

a) a fragment of a polypeptide comprising the amino acid sequence of SEQ ID Nos. 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 112, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID Nos. 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 12, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164;

b) a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID Nos. 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 112, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207192, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207223, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-34, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207189, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207221, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule comprising SEQ ID Nos. 1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 70, 72, 74, 84, 86, 88, 90, 96, 98, 100, 109, 111, 113, 122, 124, 126, 132, 134, 136, 142, 144, 146, 156, 159, 161, or 163, or a complement thereof under stringent conditions; and

c) a polypeptide which is encoded by a nucleic acid molecule comprising a nucleotide sequence which is at least 55% identical to a nucleic acid comprising the nucleotide sequence of SEQ ID Nos. 1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 70, 72, 74, 84, 86, 88, 90, 96, 98, 100, 109, 111, 113, 122, 124, 126, 132, 134, 136, 142, 144, 146, 156, 159, 161, or 163, or a complement thereof.

9. The isolated polypeptide of claim 8 comprising the amino acid sequence of SEQ ID Nos. 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 112, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164.

10. The polypeptide of claim 8 further comprising heterologous amino acid sequences.

11. An antibody which selectively binds to a polypeptide of claim 8.

12. A method for producing a polypeptide selected from the group consisting of:

a) a polypeptide comprising the amino acid sequence of SEQ ID Nos. 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 112, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207192, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207223, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-34, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207189, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207221;

b) a polypeptide comprising a fragment of the amino acid sequence of SEQ ID Nos. 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 112, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207192, the amino acid sequence encoded by the cDNA insert of the

plasmid deposited with the ATCC® as Accession Number 207223, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-34, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207189, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207221, wherein the fragment comprises at least 15 contiguous amino acids of SEQ ID Nos. 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 112, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207192, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207223, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-34, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207189, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207221; and

c) a naturally occurring allelic variant of a polypeptide comprising the amino acid sequence of SEQ ID Nos. 2, 5, 8, 11, 14, 17, 20, 71, 73, 75, 85, 87, 89, 97, 99, 101, 110, 112, 114, 123, 125, 127, 133, 135, 137, 143, 145, 147, 157, 160, 162, or 164, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207192, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207223, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number PTA-34, the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207189, or the amino acid sequence encoded by the cDNA insert of the plasmid deposited with the ATCC® as Accession Number 207221, wherein the polypeptide is encoded by a nucleic acid molecule which hybridizes to a nucleic acid molecule comprising SEQ ID Nos. 1, 3, 4, 6, 7, 9, 10, 12, 13, 15, 16, 18, 19, 21, 70, 72, 74, 84, 86, 88, 90, 96, 98, 100, 109, 111, 113, 122, 124, 126, 132, 134, 136, 142, 144, 146, 156, 159, 161, or 163, or a complement thereof under stringent conditions;

comprising culturing the host cell of claim 5 under conditions in which the nucleic acid molecule is expressed.

13. A method for detecting the presence of a polypeptide of claim 8 in a sample, comprising:

a) contacting the sample with a compound which selectively binds to a polypeptide of claim 8; and

b) determining whether the compound binds to the polypeptide in the sample.

14. The method of claim 13, wherein the compound which binds to the polypeptide is an antibody.

15. A kit comprising a compound which selectively binds to a polypeptide of claim 8 and instructions for use.

16. A method for detecting the presence of a nucleic acid molecule of claim 1 in a sample, comprising the steps of:

a) contacting the sample with a nucleic acid probe or primer which selectively hybridizes to the nucleic acid molecule; and

b) determining whether the nucleic acid probe or primer binds to a nucleic acid molecule in the sample.

17. The method of claim 16, wherein the sample comprises mRNA molecules and is contacted with a nucleic acid probe.

18. A kit comprising a compound which selectively hybridizes to a nucleic acid molecule of claim 1 and instructions for use.

19. A method for identifying a compound which binds to a polypeptide of claim 8 comprising the steps of:

a) contacting a polypeptide, or a cell expressing a polypeptide of claim 8 with a test compound; and

b) determining whether the polypeptide binds to the test compound.

20. The method of claim 19, wherein the binding of the test compound to the polypeptide is detected by a method selected from the group consisting of:

a) detection of binding by direct detecting of test compound/polypeptide binding;

b) detection of binding using a competition binding assay;

c) detection of binding using an assay for TANGO 204, TANGO 206-, TANGO 209-, or A236-mediated signal transduction.

21. A method for modulating the activity of a polypeptide of claim 8 comprising contacting a polypeptide or a cell expressing a polypeptide of claim 8 with a compound which binds to the polypeptide in a sufficient concentration to modulate the activity of the polypeptide.

22. A method for identifying a compound which modulates the activity of a polypeptide of claim 8, comprising:

a) contacting a polypeptide of claim 8 with a test compound; and

b) determining the effect of the test compound on the activity of the polypeptide to thereby identify a compound which modulates the activity of the polypeptide.

```
GTCGACCCACGCGTCCGCTCCCGGGCCGCGGCCAGCCGAGGATCAGAGCGCTTGCCCCGGGGGTCTGTCGCGTCGCAGG      79

                      M   R   T   L   W   M   A   L   C   A   L   S   R   L   W      15
CGCCGCCGGAGCCCTGGCC ATG AGG ACC CTG TGG ATG GCG CTG TGC GCG CTG TCG CGG CTG TGG      143

 P   G   A   Q   A   G   C   A   E   A   G   R   C   C   P   G   R   D   P   A      35
CCC GGG GCC CAG GCC GGC TGC GCC GAG GCC GGG CGC TGC TGT CCC GGC CGG GAC CCC GCC      203

 C   F   A   R   G   W   R   L   D   R   V   Y   G   T   C   F   C   D   Q   A      55
TGC TTC GCC CGC GGC TGG AGG CTG GAC AGG GTC TAC GGG ACG TGT TTC TGC GAC CAA GCC      263

 C   R   F   T   G   D   C   C   F   D   Y   D   R   A   C   P   A   R   P   C      75
TGT CGC TTC ACC GGG GAC TGC TGC TTC GAC TAC GAC AGG GCG TGC CCA GCT CGC CCG TGC      323

 F   V   G   E   W   S   P   W   S   G   C   A   D   Q   C   K   P   T   T   R      95
TTC GTG GGG GAA TGG AGC CCC TGG AGT GGT TGT GCA GAC CAG TGC AAG CCT ACA ACC CGT      383

 V   R   R   R   S   V   Q   Q   E   P   Q   N   G   G   A   P   C   P   P   L     115
GTG CGG AGG CGC TCG GTG CAG CAG GAG CCT CAG AAC GGC GGG GCG CCC TGC CCA CCC CTG      443

 E   E   R   A   G   C   L   E   Y   S   T   P   Q   G   Q   D   C   G   H   T     135
GAA GAG AGA GCT GGC TGC CTG GAG TAC TCC ACC CCG CAG GGC CAG GAC TGC GGG CAC ACC      503

 Y   V   P   A   F   I   T   T   S   A   F   N   K   E   R   T   R   Q   A   T     155
TAT GTT CCT GCC TTT ATA ACT ACC TCC GCA TTC AAC AAG GAG AGA ACA CGA CAA GCT ACG      563

 S   P   H   W   S   T   H   T   E   D   A   G   Y   C   M   E   F   K   T   E     175
TCT CCA CAC TGG TCT ACA CAC ACA GAG GAT GCT GGA TAC TGT ATG GAG TTT AAG ACA GAG      623
```

Fig. 1A

```
 S  L  T  P  H  C  A  L  E  N  W  P  L  T  R  W  M  Q  Y  L      195
TCC TTG ACT CCT CAC TGT GCT CTG GAA AAC TGG CCC TTG ACT AGA TGG ATG CAG TAT CTC  683

 R  E  G  Y  T  V  C  V  D  C  Q  P  P  A  M  N  S  V  S  L      215
CGA GAG GGA TAC ACG GTG TGT GTG GAT TGT CAG CCT CCA GCT ATG AAC TCT GTG AGC CTT  743

 R  C  S  G  D  G  L  D  S  D  G  N  Q  T  L  H  W  Q  A  I      235
CGT TGT TCT GGA GAT GGC CTG GAC TCC GAT GGA AAT CAG ACT CTC CAT TGG CAA GCA ATT  803

 G  N  P  R  C  Q  G  T  W  K  K  V  R  R  V  D  Q  C  S  C      255
GGT AAT CCT CGG TGT CAA GGA ACT TGG AAA AAA GTT CGG CGA GTA GAC CAG TGT TCT TGT  863

 P  A  V  H  S  F  I  F  I  *                                    265
CCA GCT GTT CAC AGT TTT ATT TTT ATA TAG                          893

ATGGTGATATAAATATTTCCAAATGCATTTGTAAACATTCTAAATATTCTCAAGTCATGTTCAATGTTTCCTAAACCTT  972

CAATTTTGGCCAAAGTCCCCAAACACATCATTGCCACACTCTGAAGTAGAGAAAGAAAATTTAGGGGCCAGTTCTCAAG 1051

GAACACAGGTCCTTTATTTTTATTTTAACTAAGTTGAAGACCCACTCAAAAAGCTCTTGTGGTTTTATGTTCTTGACCT 1130
```

Fig. 1B

TTCAACTGGAGTCCTCTCATTCAGCAGGTGGCCCGTGAGACACAGAATACATGTCTGTTTGCTAAAGTAAAATTACTGT 1209

AACTCAGTCCGATTATTGGTGACGGAAGTGTCATTTAAGGGGATCTATGTTTTGAACCTTGCAGTCTATTTTATAATCT 1288

TTTAAAAGTTTCTCACTTTAAAACATGTATACAGGTATGCACACTCTACACATATATTTCCATATATTTAATATTCCAT 1367

AAATTCTGAAATAATTTCAAGCAAATTATCACAAATAATTTTTCCACAGGGCAAATTATTTAAAATTTTAGTAAGCATT 1446

CTGTAATGAAAAACCAAGTTATACTAAAAACATTTTTTGAAGAAGAAATTTTTTTTTTGCTTAAGAAGTGAAAGGATCA 1525

AATGCTCATTGTAGGTCAATGGGAGTTCTTTTAAATGTTATAATTTTCATGAAAAGAAATGTTGACACCAGTGAATGAA 1604

ACAATTGCTTTCATTCTGAAAATTCTACCACCATTTGCATCTAAGATTATTTCCAAGGCTTAAAGCCTGAAGCTGAATA 1683

AAATAATCTTTCAGAGTCCAGCTTCAAGTTTAGTTGATGTAAGCTCACTATTTTTTTCCTACCGCATGCATTTTCTAAT 1762

GTTTGGGGTGGATGGTGTGTCGGTTATGGAAGGCATAGACGTCATTACAGGTGCTACGATCTCACACACACACAAGGAA 1841

ATGTTAGTCTCCTTATTTTATGATTGGAAAATCAATGACCTAGAGGCAAAATGGCATGTTTAAGGACCTGGGATGACAA 1920

GTCATTCTGCAGTCAGCCACAGAGCCAAATTTGGACTCCTCAACCAGAACTCCATGAAAAGCCTGACTTTGCCAAACAC 1999

TGTGCTGGAAAAGCTAAGCCCCTTTCATTTGTGAAGTAAATTTTAAATTCAAGATATTTAGTTTAGAGAATTGAGTCTT 2078

GAGATGTAAACTACATGAGATTTCTTTGGTTTCAATTGAATAATATTCACTAACAAATGATTTACTAAAATACGTATTT 2157

CTTGGTCCTTATCATGTAATGACAGATTCACAACAGCAATAAGGATGGAGATTTCCCCAATAATTAATAACACCGAGAG 2236

Fig. 1C

```
TAGCAATATTTTTTACTGTATTTTCATTTAATTGTCAAAGTTGTTTTTGTGGAAAATTAAATTCTTCAGAAATTAATAA  2315

TAAAATATGTGAGTCTCATCAAGGCTAGCAACTTTAAGTCAGTATTAACATTTTCATACTGTTTCTAACATTTATTAAG  2394

CTAGATTTTTGAAGCTATGTCTGGGTAATTTCCAATGTTACTTAAAAGTTTTTTTTTCTTTGAGACAGAGTCTCACTCT  2473

GTTGCCCAGGCTGGAGTGCAGTGGCCCCATCTCGGCTCACTGCAAGCTCCGCCTCCCGGGTTCACGCCATTCTTCTGCC  2552

TCAGCCTCCCAAGTAGCTGGGACTACAGGCGCCGTCATCACGCCTGGCTAATTTTTTGTTTTTTTTTTTTAGTAGAGAC  2631

AGGGTTTCACTGTGTTAGCCAGGGTGGTCTCGATCTCCTGACCTCGTGATCCACCCACCTCGGCCTCCCAAGGTGCTGG  2710

GATTATGGGCGTGAGCCACTGCGCCTGGCCTGTTACTTAAAATTTTTATACTAGTAAATAAACTTTTTGTTCACATTTA  2789

TAAAAGCTCTATATTTGTTTGACTAGAAAATAACCCAAATTTTCTGGACCAGTCAGAGAATGATAAAATTTCACCAAAA  2868

ATAAAACAAAAAAATCCTTCAGAACTGATAGGTATAGGTACAACAACTACTAATAGCAATAAAGCAATGGTGCTTTTAT  2947

AGTTCCGATATGTGTTATAACATTTAACATAATTGAGGTGTATAAAATAAATTTTTCACCTTATGCATTTAAAAAAAAA  3026

AAAAAAAAAAAAAAAAAAAAAGGGCGGCCGC                                                   3057
```

Fig. 1D

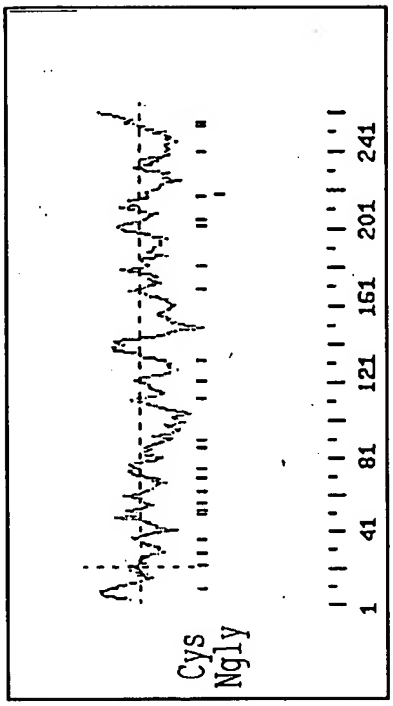

Fig. 2

```
CONCENSUS       *->dqwSC..kGfRCgEgfnag.............lkkCrCddlCksygdCC
                   +q  C ++G RC  g++ +  ++ + ++    +C Cd +C   gdCC
TANGO 204  18      AQAGCaeAG-RCCPGRDPAcfargwrldrvYGTCFCDQACRFTGDCC 63

                tDYeevCkgevs<-*
                 DY++ C +
           64   FDYDRACPARPC                                     75
```

Fig. 3

```
CONCENSUS      *->spWseWSpCSVTCGkGirtRqRtcnspaPqkkggkpCtgdaqeEtea
                 ++Ws+WS C   C   +r+R+R+++++ Pq  gg pC+ +++e  +a
TANGO 204  78  GEWSPWSGCADQCKPTTRVRRRSVQQE-PQ-NGGAPCP-PLEE--RA 119
                 CdmmdkC<-*
                       C
               120 G-----C                                    121
```

Fig. 4

```
GTCGACCCACGCGTCCGCGGACGCGTGGGTGCCCGAGGCTCTAGCTGGGTCCGTTCGCAAGTGTCGCGGCAGCTCTGGC        79

   M   K   T   L   W   M   V   L   C   A   L   A   R   L   W   P   G   A   L         19
C ATG AAG ACC CTG TGG ATG GTG CTG TGC GCT CTG GCG CGG CTG TGG CCC GGG GCC CTG         137

 A   G   C   A   E   A   G   R   C   C   P   G   R   D   P   A   C   F   A   R       39
GCT GGC TGT GCA GAG GCC GGC CGC TGC TGT CCT GGC CGG GAC CCA GCC TGC TTC GCC CGC       197

 G   W   R   L   D   R   V   Y   G   T   C   F   C   D   Q   A   C   R   L   T       59
GGC TGG AGG CTG GAC AGG GTC TAT GGA ACA TGC TTC TGT GAC CAA GCC TGC CGT CTC ACC       257

 G   D   C   C   F   D   Y   D   R   A   C   P   A   R   P   C   F   V   G   E       79
GGG GAC TGC TGC TTC GAC TAC GAC AGG GCG TGC CCA GCT CGC CCT TGC TTT GTG GGA GAA       317

 W   S   P   W   S   G   C   A   G   Q   C   Q   P   T   T   R   V   R   R   R       99
TGG AGT CCC TGG AGT GGC TGC GCA GGT CAG TGC CAG CCC ACC ACG CGT GTG CGC AGG CGT       377

 S   V   R   Q   E   P   L   N   G   G   A   P   C   P   P   L   E   E   R   A      119
TCA GTG CGT CAG GAA CCA CTA AAC GGA GGG GCG CCC TGC CCA CCC CTG GAA GAG AGA GCT       437

 G   C   L   E   Y   S   S   S   Q   S   Q   D   C   G   H   S   F   V   P   A      139
GGC TGC CTG GAG TAC TCG TCA TCA CAG AGC CAG GAC TGC GGG CAC TCC TTT GTT CCT GCC       497

 F   I   T   S   S   V   F   N   K   K   R   I   I   Q   A   V   S   P   Q   W      159
TTC ATA ACT AGC TCT GTC TTC AAC AAG AAG AGA ATA ATA CAA GCA GTA TCT CCA CAG TGG       557

 S   T   H   T   K   D   A   G   Y   C   M   E   F   K   T   E   S   L   T   P      179
TCT ACA CAC ACT AAG GAT GCT GGA TAC TGT ATG GAG TTC AAG ACA GAG TCC CTG ACT CCT       617
```

Fig. 5A

```
 H   C   A   L   V   N   S   P   L   T   R   W   M   Q   Y   L   R   E   G   Y      199
CAC TGT GCT CTG GTC AAC AGT CCC CTG ACT CGA TGG ATG CAG TAT CTC CGT GAG GGA TAC      677

 T   V   C   V   D   C   Q   P   P   A   M   N   S   V   S   L   R   C   S   G      219
ACA GTG TGT GTA GAC TGT CAG CCA CCA GCT ATG AAC TCT GTG AGC CTA CGT TGT TCT GGA      737

 D   G   L   D   S   D   G   N   Q   T   L   R   W   Q   A   I   G   N   P   R      239
GAT GGT CTG GAC TCT GAT GGA AAT CAG ACT CTT CGC TGG CAA GCC ATT GGC AAT CCT CGA      797

 C   Q   G   T   W   K   K   V   R   R   V   E   Q   C   S   C   P   D   V   H      259
TGT CAG GGA ACC TGG AAA AAA GTG CGT CGA GTA GAA CAG TGC TCA TGT CCA GAT GTG CAC      857

 R   F   I   F   I   *                                                              265
CGC TTC ATT TTT ATA TAG                                                           875

AGCACCTGGGAGCGACTTCCGAGCATGCGTGTGAATGGGCCGAGTATTGACAAGCTGCATTCAAAGCTTCCGAGGCCTC      954

CAAAAGCTGCGCCAGGGACCTGGACTCTCGCATTGCTACGCTTTGCCACTGAGAAAGAACATTTAGGGATAGTTCTCAA      1033

GGTAGACCACAGTCTCTTGGTTTTTATTTTTAATGAGTTAAGGACTCGTACTAAAAGCTCCTATGGCCTGTAGACTTTT      1112

ACCTTACATTTGGTCACAGCATACATGTCTATACTAGAATCACTACCATGTGCTCCAGTTTGGGGCAAATTGTTGTTTA      1191

AGGCACCTATGTTTTCTCTTTCTCTCACCCTCCCTCTTTTAAAAATTAAAGGTTCTCATTATATAGCCTGAAGTTCAAA      1270

AAAAAAAAAAAAAAGGGCGGCCGC                                                          1294
```

Fig. 5B

```
          10        20        30        40        50        60        70
MURINE ATGAGGACCCTGTGGATGGCGCTGTGCGCGCTGTCGCGGCTGTGGCCCCGGGGCCCAGGCCGGCTGCGCCG
       ::::.::::::::::::::: ::::::::: ::: :::::::::::::::::::::.::: ::::: :: : :
HUMAN  ATGAAGACCCTGTGGATGGTGCTGTGCGCTCTGGCGCGGCTGTGGCCCCGGGGCCCTGGCTGGCTGTGCAG
          10        20        30        40        50        60        70

          80        90       100       110       120       130       140
AGGCCGGGCGCTGCTGTCCCGGCCGGGACCCCCGCCTGCTTCGCCCGCGGCTGGAGGCTGGACAGGGTCTA
::::::: :::::::::::: :::::::::::: :::::::::::::::::::::::::::::::::::::
AGGCCGGCCGCTGCTGTCCTGGCCGGGACCCAGCCTGCTTCGCCCGCGGCTGGAGGCTGGACAGGGTCTA
          80        90       100       110       120       130       140

         150       160       170       180       190       200       210
CGGGACGTGTTTCTGCGACCAAGCCTGTCGCTTCACCGGGGACTGCTGCTTCGACTACGACAGGGCGTGC
 ::.::.:: ::::: :::::::::::: :: ::::::::::::::::::::::::::::::::::::::::
TGGAACATGCTTCTGTGACCAAGCCTGCCGTCTCACCGGGGACTGCTGCTTCGACTACGACAGGGCGTGC
         150       160       170       180       190       200       210

         220       230       240       250       260       270       280
CCAGCTCGCCCGTGCTTCGTGGGGGAATGGAGCCCCTGGAGTGGTTGTGCAGACCAGTGCAAGCCTACAA
:::::::::: ::::: :::::.::::::::: :::::::::::: :: ::::. :::::: :::: :: : :
CCAGCTCGCCCTTGCTTTGTGGGAGAATGGAGTCCCTGGAGTGGCTGCGCAGGTCAGTGCCAGCCCACCA
         220       230       240       250       260       270       280

         290       300       310       320       330       340       350
CCCGTGTGCGGAGGCGCTCGGTGCAGCAGGAGCCTCAGAACGGCGGGGCGCCCTGCCCACCCCTGGAAGA
: :::::::: ::::: ::.:::::. :::::.::..:.:::: ::::::::::::::::::::::::::::
CGCGTGTGCGCAGGCGTTCAGTGCGTCAGGAACCACTAAACGGAGGGGCGCCCTGCCCACCCCTGGAAGA
         290       300       310       320       330       340       350
```

Fig. 6A

```
      360       370       380       390       400       410       420
GAGAGCTGGCTGCCTGGAGTACTCCACCCCGCAGGGCCAGGACTGCGGGCACACCTATGTTCCTGCCTTT
:::::::::::::::::::::::: .: :.:::.:::::::::::::::::.:::.:::::::::::::
GAGAGCTGGCTGCCTGGAGTACTCGTCATCACAGAGCCAGGACTGCGGGCACTCCTTTGTTCCTGCCTTC
      360       370       380       390       400       410       420

      430       440       450       460       470       480       490
ATAACTACCTCCGCATTCAACAAGGAGAGAACACGACAAGCTACGTCTCCACACTGGTCTACACACACAG
::::::: ::: :  :::::::::.:::::: :  ::::::.. .:::::::: ::::::::::::::..
ATAACTAGCTCTGTCTTCAACAAGAAGAGAATAATACAAGCAGTATCTCCACAGTGGTCTACACACACTA
      430       440       450       460       470       480       490

      500       510       520       530       540       550       560
AGGATGCTGGATACTGTATGGAGTTTAAGACAGAGTCCTTGACTCCTCACTGTGCTCTGGAAAACTGGCC
:::::::::::::::::::::::: :::::::::::: :::::::::::::::::::. ::::.: ::
AGGATGCTGGATACTGTATGGAGTTCAAGACAGAGTCCCTGACTCCTCACTGTGCTCTGGTCAACAGTCC
      500       510       520       530       540       550       560

      570       580       590       600       610       620       630
CTTGACTAGATGGATGCAGTATCTCCGAGAGGGATACACGGTGTGTGTGGATTGTCAGCCTCCAGCTATG
: :::: :::::::::::::::::::.:::::::::::.:: ::::::::.:::::::::
CCTGACTCGATGGATGCAGTATCTCCGTGAGGGATACACAGTGTGTGTAGACTGTCAGCCACCAGCTATG
      570       580       590       600       610       620       630

      640       650       660       670       680       690       700
AACTCTGTGAGCCTTCGTTGTTCTGGAGATGGCCTGGACTCCGATGGAAATCAGACTCTCCATTGGCAAG
:::::::::::::::::.:::::::::::::::: :::::::: :::::::::::::::::: :. :::::::
AACTCTGTGAGCCTACGTTGTTCTGGAGATGGTCTGGACTCTGATGGAAATCAGACTCTTCGCTGGCAAG
      640       650       660       670       680       690       700
```

Fig. 6B

```
     710       720       730       740       750       760       770
CAATTGGTAATCCTCGGTGTCAAGGAACTTGGAAAAAAGTTCGGCGAGTAGACCAGTGTTCTTGTCCAGC
: ::::: :::::::::.:::::.::::: ::::::::::: :: ::::::::: ::::: ::.::::::
CCATTGGCAATCCTCGATGTCAGGGAACCTGGAAAAAAGTGCGTCGAGTAGAACAGTGCTCATGTCCAGA
     710       720       730       740       750       760       770

     780       790
TGTTCACAGTTTTATTTTTATA
::: ::: : :: :::::::::
TGTGCACCGCTTCATTTTTATA
     780       790
```

Fig. 6C

```
                10        20        30        40        50        60        70
HUMAN   MRTLWMALCALSRLWPGAQAGCAEAGRCCPGRDPACFARGWRLDRVYGTCFCDQACRFTGDCCFDYDRAC
        :.:::::.::::.:::::: :::::::::::::::::::::::::::::::::::.:::::::::::::
MURINE  MKTLWMVLCALARLWPGALAGCAEAGRCCPGRDPACFARGWRLDRVYGTCFCDQACRLTGDCCFDYDRAC
                10        20        30        40        50        60        70

                80        90       100       110       120       130       140
        PARPCFVGEWSPWSGCADQCKPTTRVRRRSVQQEPQNGGAPCPPLEERAGCLEYSTPQGQDCGHTYVPAF
        :::::::::::::::::::.::.:::::::::::::: :::::::::::::::::::..:.:::::..::::
        PARPCFVGEWSPWSGCAGQCQPTTRVRRRSVRQEPLNGGAPCPPLEERAGCLEYSSSQSQDCGHSFVPAF
                80        90       100       110       120       130       140

               150       160       170       180       190       200       210
        ITTSAFNKERTRQATSPHWSTHTEDAGYCMEFKTESLTPHCALENWPLTRWMQYLREGYTVCVDCQPPAM
        :::.::::: :. ::.::.:::::: :::::::::::::::::::: : :::::::::::::::::::::
        ITSSVFNKKRIIQAVSPQWSTHTKDAGYCMEFKTESLTPHCALVNSPLTRWMQYLREGYTVCVDCQPPAM
               150       160       170       180       190       200       210

               220       230       240       250       260
        NSVSLRCSGDGLDSDGNQTLHWQAIGNPRCQGTWKKVRRVDQCSCPAVHSFIFI
        :::::::::::::::::::::::::.::::::::::::::::.:::::.:: ::::
        NSVSLRCSGDGLDSDGNQTLRWQAIGNPRCQGTWKKVRRVEQCSCPDVHRFIFI
               220       230       240       250       260
```

Fig. 7

```
GTCGACCCACGCGTCCGAGACGGCCCACGGCGCCCGCGGGCTGGGGCGGTCGCTTCTTCCTTCTCCGTGGCCTACGAGG        79

                      M   A   P   W   P   P   K   G   L   V   P   A   V   L   W        15
GTCCCCAGCCTGGGTAAAG ATG GCC CCA TGG CCC CCG AAG GGC CTA GTC CCA GCT GTG CTC TGG        143

 G   L   S   L   F   L   N   L   P   G   P   I   W   L   Q   P   S   P   P   P        35
GGC CTC AGC CTC TTC CTC AAC CTC CCA GGA CCT ATC TGG CTC CAG CCC TCT CCA CCT CCC        203

 Q   S   S   P   P   S   Q   P   H   P   C   H   T   C   R   G   L   V   D   S        55
CAG TCT TCT CCC CCG TCT CAG CCC CAT CCG TGT CAT ACC TGC CGG GGA CTG GTT GAC AGC        263

 F   N   K   G   L   E   R   T   I   R   D   N   F   G   G   G   N   T   A   W        75
TTT AAC AAG GGC CTG GAG AGA ACC ATC CGG GAC AAC TTT GGA GGT GGA AAC ACT GCC TGG        323

 E   E   E   N   L   S   K   Y   K   D   S   E   T   R   L   V   E   V   L   E        95
GAG GAA GAG AAT TTG TCC AAA TAC AAA GAC AGT GAG ACC CGC CTG GTA GAG GTG CTG GAG        383

 G   V   C   S   K   S   D   F   E   C   H   R   L   L   E   L   S   E   E   L       115
GGT GTG TGC AGC AAG TCA GAC TTC GAG TGC CAC CGC CTG CTG GAG CTG AGT GAG GAG CTG        443

 V   E   S   W   W   F   H   K   Q   Q   E   A   P   D   L   F   Q   W   L   C       135
GTG GAG AGC TGG TGG TTT CAC AAG CAG CAG GAG GCC CCG GAC CTC TTC CAG TGG CTG TGC        503

 S   D   S   L   K   L   C   C   P   A   G   T   F   G   P   S   C   L   P   C       155
TCA GAT TCC CTG AAG CTC TGC TGC CCC GCA GGC ACC TTC GGG CCC TCC TGC CTT CCC TGT        563

 P   G   G   T   E   R   P   C   G   G   Y   G   Q   C   E   G   E   G   T   R       175
CCT GGG GGA ACA GAG AGG CCC TGC GGT GGC TAC GGG CAG TGT GAA GGA GAA GGG ACA CGA        623
```

Fig. 8A

```
 G  G  S  G  H  C  D  C  Q  A  G  Y  G  G  E  A  C  G  Q  C      195
GGG GGC AGC GGG CAC TGT GAC TGC CAA GCC GGC TAC GGG GGT GAG GCC TGT GGC CAG TGT      683

 G  L  G  Y  F  E  A  E  R  N  A  S  H  L  V  C  S  A  C  F      215
GGC CTT GGC TAC TTT GAG GCA GAA CGC AAC GCC AGC CAT CTG GTA TGT TCG GCT TGT TTT      743

 G  P  C  A  R  C  S  G  P  E  E  S  N  C  L  Q  C  K  K  G      235
GGC CCC TGT GCC CGA TGC TCA GGA CCT GAG GAA TCA AAC TGT TTG CAA TGC AAG AAG GGC      803

 W  A  L  H  H  L  K  C  V  D  I  D  E  C  G  T  E  G  A  N      255
TGG GCC CTG CAT CAC CTC AAG TGT GTA GAC ATT GAT GAG TGT GGC ACA GAG GGA GCC AAC      863

 C  G  A  D  Q  F  C  V  N  T  E  G  S  Y  E  C  R  D  C  A      275
TGT GGA GCT GAC CAA TTC TGC GTG AAC ACT GAG GGC TCC TAT GAG TGC CGA GAC TGT GCC      923

 K  A  C  L  G  C  M  G  A  G  P  G  R  C  K  K  C  S  P  G      295
AAG GCC TGC CTA GGC TGC ATG GGG GCA GGG CCA GGT CGC TGT AAG AAG TGT AGC CCT GGC      983

 Y  Q  Q  V  G  S  K  C  L  D  V  D  E  C  E  T  E  V  C  P      315
TAT CAG CAG GTG GGC TCC AAG TGT CTC GAT GTG GAT GAG TGT GAG ACA GAG GTG TGT CCG     1043

 G  E  N  K  Q  C  E  N  T  E  G  G  Y  R  C  I  C  A  E  G      335
GGA GAG AAC AAG CAG TGT GAA AAC ACC GAG GGC GGT TAT CGC TGC ATC TGT GCC GAG GGC     1103

 Y  K  Q  M  E  G  I  C  V  K  E  Q  I  P  E  S  A  G  F  F      355
TAC AAG CAG ATG GAA GGC ATC TGT GTG AAG GAG CAG ATC CCA GAG TCA GCA GGC TTC TTC     1163

 S  E  M  T  E  D  E  L  V  V  L  Q  Q  M  F  F  G  I  I  I      375
TCA GAG ATG ACA GAA GAC GAG TTG GTG GTG CTG CAG CAG ATG TTC TTT GGC ATC ATC ATC     1223
```

Fig. 8B

```
 C   A   L   A   T   L   A   A   K   G   D   L   V   F   T   A   I   F   I   G      395
TGT GCA CTG GCC ACG CTG GCT GCT AAG GGC GAC TTG GTG TTC ACC GCC ATC TTC ATT GGG      1283

 A   V   A   A   M   T   G   Y   W   L   S   E   R   S   D   R   V   L   E   G      415
GCT GTG GCG GCC ATG ACT GGC TAC TGG TTG TCA GAG CGC AGT GAC CGT GTG CTG GAG GGC      1343

 F   I   K   G   R   *                                                              421
TTC ATC AAG GGC AGA TAA                                                            1361

TCGCGGCCACCACCTGTAGGACCTCCTCCCACCCACGCTGCCCCCAGAGCTTGGGCTGCCCTCCTGCTGGACACTCAGG     1440

ACAGCTTGGTTTATTTTTGAGAGTGGGGTAAGCACCCCTACCTGCCTTACAGAGCAGCCCAGGTACCCAGGCCCGGGCA     1519

GACAAGGCCCCTGGGGTAAAAAGTAGCCCTGAAGGTGGATACCATGAGCTCTTCACCTGGCGGGGACTGGCAGGCTTCA     1598

CAATGTGTGAATTTCAAAAGTTTTTCCTTAATGGTGGCTGCTAGAGCTTTGGCCCCTGCTTAGGATTAGGTGGTCCTCA     1677

CAGGGGTGGGGCCATCACAGCTCCCTCCTGCCAGCTGCATGCTGCCAGTTCCTGTTCTGTGTTCACCACATCCCCACAC     1756

CCCATTGCCACTTATTTATTCATCTCAGGAAATAAAGAAAGGTCTTGGAAAGTTAAAAAAAAAAAAAAAAAAAGGGCG     1835

GCCGC                                                                              1840
```

Fig. 8C

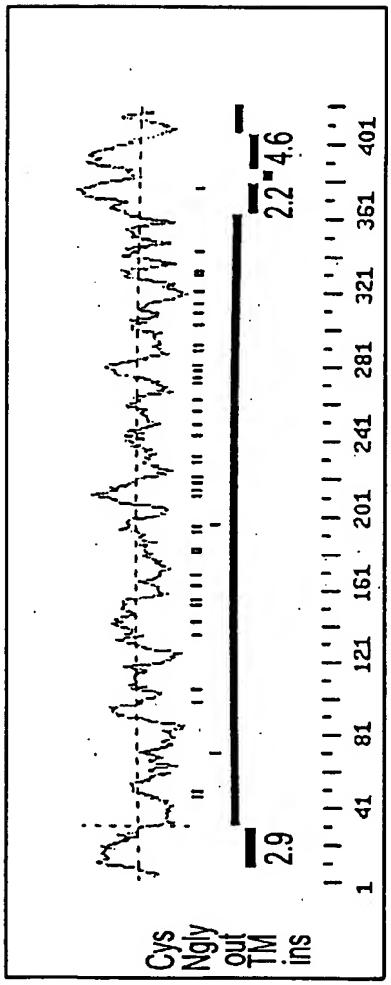
Cys
Ngly
out
TM
ins
2.9
2.2
4.6
1
41
81
121
161
201
241
281
321
361
401

Fig. 9

```
CONCENSUS     *->CdCnphGslsddtCdsddelfgeetGqClkCkpnvtGrrCdrCkpGy
                 +C + G+++          +G+C+ C+ +++G+ C +C  Gy
TANGO 206 168    -QCEGEGTRG----------GSGHCD-CQAGYGGEACGQCGLGY 199

                 yglpsg.dpgqgC<-*
                 +  ++++ ++  C
              200 F-EAERnASHLVC                              211
```

Fig. 10

```
GTCGACCCACGCGTCCGCCTCTCTCTAATTTTGGCCCCTAATAACTCACTGAGACCCTAAAACAATGGGCTGACGCTTA          79

TTCTTTTGCCCGCGATCCCGGCCACCTCCCCCTACGGTAACGGAGAAATAATTTCTCTGCTGGAATCGTAGATTAGGCT         158

CTTTCACCCCTTCCAGCGCCTCTGCTTCCACTACCCCAGTTTATCCTCGTTTTGTAATTACGTAGGTCTGGGGCTCTGA         237

GCCCCAAGGCGCCTCGCTTTCCTCCCTTACAGCCTGGCTTCTCCAGTGAACCCTCCCCCTGCCCTCTATCTGCACGTCC         316

                    M   A   P   L   P   P   R   G   L   V   P   S   L   L   W   C          16
CCAGCCTGGGAAAAG ATG GCT CCA CTG CCC CCA AGG GGC CTG GTC CCA TCT CTG CTC TGG TGC          379

 L   S   L   F   L   S   L   P   G   P   V   W   L   Q   P   S   P   P   P   H          36
TTG AGC CTG TTT CTG AGC CTC CCA GGA CCT GTC TGG CTC CAA CCC TCT CCT CCT CCC CAT          439

 P   S   P   R   A   E   P   H   P   C   H   T   C   R   A   L   V   D   N   F          56
CCT TCT CCC CGA GCT GAG CCC CAT CCG TGT CAT ACC TGC CGG GCA CTG GTG GAC AAC TTC          499

 N   K   G   L   E   R   T   I   R   D   N   F   G   G   G   N   T   A   W   E          76
AAC AAG GGC CTG GAG AGA ACC ATC CGG GAC AAC TTC GGG GGT GGA AAC ACG GCC TGG GAG          559

 E   E   K   L   S   K   Y   K   D   S   E   T   R   L   V   E   V   L   E   G          96
GAA GAG AAG TTG TCC AAA TAC AAA GAC AGT GAG ACC CGC CTG GTG GAG GTG CTG GAG GGC          619

 V   C   S   R   S   D   F   E   C   H   R   L   L   E   L   S   E   E   L   V         116
GTG TGC AGC AGG TCA GAC TTC GAG TGC CAC CGC CTG CTC GAG CTG AGC GAG GAG CTG GTG          679

 E   N   W   W   F   H   R   Q   Q   E   A   P   D   L   F   Q   W   L   C   S         136
GAA AAC TGG TGG TTT CAC AGG CAG CAG GAA GCC CCC GAC CTC TTC CAG TGG CTC TGT TCC          739
```

Fig. 11A

```
  D   S   L   K   L   C   C   P   S   G   G   T   F   G   P   S   C   L   P   C   P      156
GAT TCC CTG AAG CTC TGC TGC CCC TCT GGC ACC TTT GGG CCC TCC TGC CTG CCA TGT CCT      799

  G   G   T   E   R   P   C   G   G   Y   G   Q   C   E   G   E   G   T   R   G      176
GGG GGC ACA GAG AGG CCC TGC GGT GGC TAC GGG CAG TGT GAA GGG GAA GGG ACT CGA GGG      859

  G   S   G   H   C   D   C   Q   A   G   Y   G   G   E   A   C   G   Q   C   G      196
GGC AGC GGG CAC TGT GAC TGC CAA GCC GGC TAT GGG GGC GAG GCC TGT GGC CAG TGT GGC      919

  L   G   Y   F   E   A   E   R   N   S   S   H   L   V   C   S   A   C   F   G      216
CTT GGC TAC TTT GAG GCG GAG CGC AAC AGC AGC CAT CTG GTA TGT TCG GCG TGT TTT GGT      979

  P   C   A   R   C   T   G   P   E   E   S   H   C   L   Q   C   K   K   G   W      236
CCT TGT GCC CGC TGC ACG GGA CCT GAG GAA TCC CAT TGT CTG CAG TGC AAG AAA GGC TGG     1039

  A   L   H   H   L   K   C   V   D   I   D   E   C   G   T   E   Q   A   T   C      256
GCC CTG CAT CAC CTC AAG TGT GTA GAC ATC GAT GAG TGT GGT ACA GAG CAA GCC ACC TGT     1099

  G   A   D   Q   F   C   V   N   T   E   G   S   Y   E   C   R   D   C   A   K      276
GGA GCT GAC CAG TTC TGT GTG AAC ACA GAA GGA TCC TAC GAG TGC CGA GAC TGT GCA AAG     1159
```

Fig. 11B

```
 A   C   L   G   C   M   G   A   G   P   G   R   C   K   K   C   S   R   G   Y      296
GCC TGC CTG GGC TGT ATG GGA GCA GGG CCA GGG CGC TGC AAA AAG TGC AGC CGT GGC TAC      1219

 Q   Q   V   G   S   K   C   L   D   V   D   E   C   E   T   V   V   C   P   G      316
CAG CAG GTG GGC TCC AAG TGC CTA GAT GTG GAT GAG TGT GAG ACT GTG GTG TGT CCA GGA      1279

 E   N   E   K   C   E   N   T   E   G   G   Y   R   C   V   C   A   E   G   Y      336
GAG AAT GAG AAG TGT GAA AAC ACG GAG GGA GGC TAC CGC TGT GTC TGT GCT GAG GGC TAC      1339

 R   Q   E   D   G   I   C   V   K   E   Q   V   P   E   S   A   G   F   F   A      356
AGA CAG GAG GAC GGC ATC TGC GTG AAG GAG CAG GTC CCG GAG TCG GCG GGC TTC TTT GCG      1399

 E   M   T   E   D   E   M   V   V   L   Q   Q   M   F   F   G   V   I   I   C      376
GAG ATG ACG GAG GAT GAA ATG GTG GTT CTG CAG CAG ATG TTC TTT GGT GTA ATC ATC TGT      1459

 A   L   A   T   L   A   A   K   G   D   L   V   F   T   A   I   F   I   G   A      396
GCG CTG GCC ACG CTT GCT GCC AAG GGT GAC TTG GTG TTC ACT GCC ATC TTC ATT GGA GCT      1519

 V   A   A   M   T   G   Y   W   L   S   E   R   S   D   R   V   L   E   G   F      416
GTG GCA GCC ATG ACT GGG TAC TGG TTG TCA GAG CGG AGT GAC CGT GTG CTG GAG GGC TTC      1579

 I   K   G   R   *                                                                  421
ATC AAG GGT AGA TAA                                                                 1594

TCCCTGCCGCCACTCACAGGATTTCCTCCCACCCAGGCTGCTCCTAGAGGTTATTTCTCTCTCCCTATGGATACCTGGG    1673

ACAGCATAGTTTCTCTTTGAGACTGGGGTAAACACCCTTTTACCTGCCTTACTGAGCAGCCCAGGACCCCAGTGGCCCC    1752
```

Fig. 11C

AGGGGGTTGAAAAAGAGGTCCTGAAGCAGATGCCACCGAGATCCTAGCCTAGGACACTGGCAGGCCTCACCGTGTATGAA 1831

TCCCAAGACAGTTTCTCCCCAGGGAGGTTGCAGGTGGGCTTTGGCCCCTGACCCGGATGAGAGGTGTCCCCTCAGGGGT 1910

GGGCCCCCATAGGCCCCTTGCCAGGTGCATGCTGCCAGCTCCAGCTGTGTATTCACCGCCCACGCCCTAGCCATCGACT 1989

TATTTATTCATCTCAGGAAATAAAGGACGGTCTTGAAAATGGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2068

AAAAAAAAAAAAAAAGGGCGGCCGC 2093

Fig. 11D

```
          10        20        30        40        50        60        70
MURINE    ATGGCCCCCATGGCCCCCGAAGGGCCTAGTCCCAGCTGTGCTCTGGGGCCTCAGCCTCTTCCTCAACCTCC
          ::::: :::  :::::::.:.:::::::.:::::: :: :::::::: :: : ::::: :: :: :.::::::
HUMAN     ATGGCTCCACTGCCCCCAAGGGCCTGGTCCCATCTCTGCTCTGGTGCTTGAGCCTGTTTCTGAGCCTCC
          10        20        30        40        50        60        70

          80        90        100       110       120       130       140
          CAGGACCTATCTGGCTCCAGCCCTCTCCACCTCCCCAGTCTTCTCCCCCGTCTCAGCCCCATCCGTGTCA
          :::::::::.:::::::::::::::.:::::::::  ::::::::: . :: ::::::::::::::::::
          CAGGACCTGTCTGGCTCCAACCCTCTCCTCCCCATCCTTCTCCCCGAGCTGAGCCCCATCCGTGTCA
          80        90        100       110       120       130       140

          150       160       170       180       190       200       210
          TACCTGCCGGGGACTGGTTGACAGCTTTAACAAGGGCCTGGAGAGAACCATCCGGGACAACTTTGGAGGT
          :::::::::: :::::: :::.::: :::::::::::::::::::::::::::::::::: ::.:::
          TACCTGCCGGGCACTGGTGGACAACTTCAACAAGGGCCTGGAGAGAACCATCCGGGACAACTTCGGGGGT
          150       160       170       180       190       200       210

          220       230       240       250       260       270       280
          GGAAACACTGCCTGGGAGGAAGAGAATTTGTCCAAATACAAAGACAGTGAGAGACCCGCCTGGTAGAGGTGC
          :::::::: :::::::::::::::::: :::::::::::::::::::::::::::::::.:::::::
          GGAAACACGGCCTGGGAGGAAGAGAAGTTGTCCAAATACAAAGACAGTGAGAGACCCGCCTGGTGGAGGTGC
          220       230       240       250       260       270       280

          290       300       310       320       330       340       350
          TGGAGGGTGTGTGCAGCAAGTCAGACTTCGAGTGCCACCGCCTGCTGGAGCTGAGTGAGGAGCTGGTGGA
          ::::::: :::::::::::.:::::::::::::::::::::::::: :::::::: :::::::::::::
          TGGAGGGCGTGTGCAGCAGGTCAGACTTCGAGTGCCACCGCCTGCTCGAGCTGAGCGAGGAGCTGGTGGA
          290       300       310       320       330       340       350
```

Fig. 12A

```
      360       370       380       390       400       410       420
GAGCTGGTGGTTTCACAAGCAGCAGGAGGCCCCGGACCTCTTCCAGTGGCTGTGCTCAGATTCCCTGAAG
 :.:::::::::::::::::.:::::::::.:::::  :::::::::::::::::: :: ::  :::::::::::
AAACTGGTGGTTTCACAGGCAGCAGGAAGCCCCCGACCTCTTCCAGTGGCTCTGTTCCGATTCCCTGAAG
       360       370       380       390       400       410       420

      430       440       450       460       470       480       490
CTCTGCTGCCCCGCAGGCACCTTCGGGCCCTCCTGCCTTCCCTGTCCTGGGGGAACAGAGAGGCCCTGCG
 ::::::::::: :.::::::::: :::::::::::::: :: :::::::::::: ::::::::::::::::::
CTCTGCTGCCCCTCTGGCACCTTTGGGCCCTCCTGCCATGTCCTGGGGGCACAGAGAGGCCCTGCG
      430       440       450       460       470       480       490

      500       510       520       530       540       550       560
GTGGCTACGGGCAGTGTGAAGGAGAAGGGACACGAGGGGGCAGCGGGCACTGTGACTGCCAAGCCGGCTA
 ::::::::::.:::::::::::::::.:::::::::::::::::::::::::::::::::::::::::::
GTGGCTACGGGCAGTGTGAAGGGGAAGGGACTCGAGGGGGCAGCGGGCACTGTGACTGCCAAGCCGGCTA
      500       510       520       530       540       550       560

      570       580       590       600       610       620       630
CGGGGGTGAGGCCTGTGGCCAGTGTGGCCTTGGCTACTTTGAGGCAGAACGCAACGCCAGCCATCTGGTA
 :::::::::::::::::::::::::::::::::::::::::.::.:::::::. ::::::::::::::::
TGGGGGCGAGGCCTGTGGCCAGTGTGGCCTTGGCTACTTTGAGGCGGAGCGCAACAGCAGCCATCTGGTA
      570       580       590       600       610       620       630

      640       650       660       670       680       690       700
TGTTCGGCTTGTTTTGGCCCCTGTGCCCGATGCTCAGGACCTGAGGAATCAAACTGTTTGCAATGCAAGA
 :::::::: :::::::: :: :::::::: :::.:.:::::::::::::::: : ::: ::::.:::::::
TGTTCGGCGTGTTTTGGTCCTTGTGCCCGCTGCACGGGACCTGAGGAATCCCCATTGTCTGCAGTGCAAGA
      640       650       660       670       680       690       700
```

Fig. 12B

```
      710       720       730       740       750       760       770
AGGGCTGGGCCCTGCATCACCTCAAGTGTGTAGACATTGATGAGTGTGGCACAGAGGGAGCCAACTGTGG
: .:::::::::::::::::::::::::::::::::: :::::::::::: :::::: .::::: ::::::
AAGGCTGGGCCCTGCATCACCTCAAGTGTGTAGACATCGATGAGTGTGGTACAGAGCAAGCCACCTGTGG
      710       720       730       740       750       760       770

      780       790       800       810       820       830       840
AGCTGACCAATTCTGCGTGAACACTGAGGGCTCCTATGAGTGCCGAGACTGTGCCAAGGCCTGCCTAGGC
:::::::::.::::: :::::::::.::.:: ::::: :::::::::::::::::: ::::::::::::.:::
AGCTGACCAGTTCTGTGTGAACACAGAAGGATCCTACGAGTGCCGAGACTGTGCAAAGGCCTGCCTGGGC
      780       790       800       810       820       830       840

      850       860       870       880       890       900       910
TGCATGGGGGCAGGGCCAGGTCGCTGTAAGAAGTGTAGCCCTGGCTATCAGCAGGTGGGCTCCAAGTGTC
:: ::::::::::::::: ::::: ::.::::: :::: ::::: :::::::::::::::::::: : :
TGTATGGGAGCAGGGCCAGGGCGCTGCAAAAAGTGCAGCCGTGGCTACCAGCAGGTGGGCTCCAAGTGCC
      850       860       870       880       890       900       910

      920       930       940       950       960       970       980
TCGATGTGGATGAGTGTGAGACAGAGGTGTGTCCGGGAGAGAACAAGCAGTGTGTGAAAACACCGAGGGCGG
: ::::::::::::::::::::::::::..::::::::::::: .:: :::::::::::::: ::::: ::
TAGATGTGGATGAGTGTGAGACTGTGGTGTGTCCAGGAGAGAATGAGAAGTGTGTGAAAACACGGAGGGAGG
      920       930       940       950       960       970       980

      990      1000      1010      1020      1030      1040      1050
TTATCGCTGCATCTGTGCCGAGGGCTACAAGCAGATGGAAGGCATCTGTGTGAAGGAGCAGATCCCAGAG
:: ::::: .:::::::: :::::::::::..::::..::: :::::::: ::::::::::::::.:::::
CTACCGCTGTGTCTGTGCTGAGGGCTACAGACAGGAGGACGGCATCTGCGTGAAGGAGCAGGTCCCGGAG
      990      1000      1010      1020      1030      1040      1050
```

Fig. 12C

```
      1060      1070      1080      1090      1100      1110      1120
TCAGCAGGCTTCTTCTCAGAGATGACAGAAGACGAGTTGGTGGTGCTGCAGCAGATGTTCTTTGGCATCA
:::.::.:::::::::  :.:::::::::.::.:: ::..:::::::: ::::::::::::::::::: .: :
TCGGCGGGCTTCTTTGCGGAGATGACGGAGGATGAAATGGTGGTTCTGCAGCAGATGTTCTTTGGTGTAA
      1060      1070      1080      1090      1100      1110      1120

      1130      1140      1150      1160      1170      1180      1190
TCATCTGTGCACTGGCCACGCTGGCTGCTAAGGGCGACTTGGTGTTCACCGCCATCTTCATTGGGGCTGT
:::::::::::.:::::::::::::: ::::: ::::: :::::::::::::: ::::::::::::::.:::::
TCATCTGTGCGCTGGCCACGCTTGCTGCCAAGGGTGACTTGGTGTTCACTGCCATCTTCATTGGAGCTGT
      1130      1140      1150      1160      1170      1180      1190

      1200      1210      1220      1230      1240      1250      1260
GGCGGCCATGACTGGCTACTGGTTGTCAGAGCGCAGTGACCGTGTGCTGGAGGGCTTCATCAAGGGCAGA
:::.::::::::::: :::::::::::::::: :::::::::::::::::::::::::::::: :::
GGCAGCCATGACTGGGTACTGGTTGTCAGAGCGGAGTGACCGTGTGCTGGAGGGCTTCATCAAGGGTAGA
      1200      1210      1220      1230      1240      1250      1260
```

Fig. 12D

```
                10        20        30        40        50        60        70
HUMAN   MAPWPPKGLVPAVLWGLSLFLNLPGPIWLQPSPPPQSSPPSQPHPCHTCRGLVDSFNKGLERTIRDNFGG
        ::: ::.::::..:: :::::.::::.::::::::..::  ..:::::::.:::.:::::::::::::::
MURINE  MAPLPPRGLVPSLLWCLSLFLSLPGPVWLQPSPPPHPSPRAEPHPCHTCRALVDNFNKGLERTIRDNFGG
                10        20        30        40        50        60        70

                80        90       100       110       120       130       140
        GNTAWEEENLSKYKDSETRLVEVLEGVCSKSDFECHRLLELSEELVESWWFHKQQEAPDLFQWLCSDSLK
        ::::::::.::::::::::::::::::::.:::::::::::::::::.::::.:::::::::::::::::
        GNTAWEEEKLSKYKDSETRLVEVLEGVCSRSDFECHRLLELSEELVENWWFHRQQEAPDLFQWLCSDSLK
                80        90       100       110       120       130       140

               150       160       170       180       190       200       210
        LCCPAGTFGPSCLPCPGGTERPCGGYGQCEGEGTRGGSGHCDCQAGYGGEACGQCGLGYFEAERNASHLV
        ::::.:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::.::::
        LCCPSGTFGPSCLPCPGGTERPCGGYGQCEGEGTRGGSGHCDCQAGYGGEACGQCGLGYFEAERNSSHLV
               150       160       170       180       190       200       210

               220       230       240       250       260       270       280
        CSACFGPCARCSGPEESNCLQCKKGWALHHLKCVDIDECGTEGANCGADQFCVNTEGSYECRDCAKACLG
        :::::::::::.:::::.:::::::::::::::::::::::: :.:::::::::::::::::::::::::
        CSACFGPCARCTGPEESHCLQCKKGWALHHLKCVDIDECGTEQATCGADQFCVNTEGSYECRDCAKACLG
               220       230       240       250       260       270       280

               290       300       310       320       330       340       350
        CMGAGPGRCKKCSPGYQQVGSKCLDVDECETEVCPGENKQCENTEGGYRCICAEGYKQMEGICVKEQIPE
        ::::::::::::: ::::::::::::::::: :::::: .::::::::::.:::::.: .::::::.::
        CMGAGPGRCKKCSRGYQQVGSKCLDVDECETVVCPGENEKCENTEGGYRCVCAEGYRQEDGICVKEQVPE
               290       300       310       320       330       340       350
```

Fig. 13A

```
     360       370       380       390       400       410       420
SAGFFSEMTEDELVVLQQMFFGIIICALATLAAKGDLVFTAIFIGAVAAMTGYWLSERSDRVLEGFIKGR
:::::.:::::::.:::::::::.:::::::::::::::::::::::::::::::::::::::::::::
SAGFFAEMTEDEMVVLQQMFFGVIICALATLAAKGDLVFTAIFIGAVAAMTGYWLSERSDRVLEGFIKGR
     360       370       380       390       400       410       420
```

Fig. 13B

```
GTCGACCCACGCGTCCGCCTCTGGGTGCCTGCAGGGGAGCTGCTCCAGCCGGGCCGCCGGGAGCGGTGGGGAGAGCATC          79

GCGGAGCCGCCCCTCCACGCGCCCGCCCAGCCGCGCTCGCCCACTGGGCTCTCCCGGCTGCAGTGCCAGGGCGCAGGAC          158

                                        M   L   L   P   Q   L   C   W   L   P   L          11
GCGGCCGATCTCCCGCTCCCGCCACCTCCGCCACC   ATG CTG CTC CCC CAG CTC TGC TGG CTG CCG CTG          226

 L   A   G   L   L   P   P   V   P   A   Q   K   F   S   A   L   T   F   L   R           31
CTC GCT GGG CTG CTC CCG CCG GTG CCC GCT CAG AAG TTC TCG GCG CTC ACG TTT TTG AGA          286

 V   D   Q   D   K   D   K   D   C   S   L   D   C   A   G   S   P   Q   K   P           51
GTG GAT CAA GAT AAA GAC AAG GAT TGT AGC TTG GAC TGT GCG GGT TCG CCC CAG AAA CCT          346

 L   C   A   S   D   G   R   T   F   L   S   R   C   E   F   Q   R   A   K   C           71
CTC TGC GCA TCT GAC GGA AGG ACC TTC CTT TCC CGT TGT GAA TTT CAA CGT GCC AAG TGC          406

 K   D   P   Q   L   E   I   A   Y   R   G   N   C   K   D   V   S   R   C   V           91
AAA GAT CCC CAG CTA GAG ATT GCA TAT CGA GGA AAC TGC AAA GAC GTG TCC AGG TGT GTG          466

 A   E   R   K   Y   T   Q   E   Q   A   R   K   E   F   Q   Q   V   F   I   P          111
GCC GAA AGG AAG TAT ACC CAG GAG CAA GCC CGG AAG GAG TTT CAG CAA GTG TTC ATT CCT          526

 E   C   N   D   D   G   T   Y   S   Q   V   Q   C   H   S   Y   T   G   Y   C          131
GAG TGC AAT GAC GAC GGC ACC TAC AGT CAG GTC CAG TGT CAC AGC TAC ACG GGA TAC TGC          586

 W   C   V   T   P   N   G   R   P   I   S   G   T   A   V   A   H   K   T   P          151
TGG TGC GTC ACG CCC AAC GGG AGG CCC ATC AGC GGC ACT GCC GTG GCC CAC AAG ACG CCC          646
```

Fig. 14A

```
R  C  P  G  S  V  N  E  K  L  P  Q  R  E  G  T  G  K  T  D      171
CGG TGC CCG GGT TCC GTA AAT GAA AAG TTA CCC CAA CGC GAA GGC ACA GGA AAA ACA GAT      706

D  A  A  A  P  A  L  E  T  Q  P  Q  G  D  E  E  D  I  A  S      191
GAT GCC GCA GCT CCA GCG TTG GAG ACT CAG CCT CAA GGA GAT GAA GAA GAT ATT GCA TCA      766

R  Y  P  T  L  W  T  E  Q  V  K  S  R  Q  N  K  T  N  K  N      211
CGT TAC CCT ACC CTT TGG ACT GAA CAG GTT AAA AGT CGG CAG AAC AAA ACC AAT AAG AAT      826

S  V  S  S  C  D  Q  E  H  Q  S  A  L  E  E  A  K  Q  P  K      231
TCA GTG TCA TCC TGT GAC CAA GAG CAC CAG TCT GCC CTG GAG GAA GCC AAG CAG CCC AAG      886

N  D  N  V  V  I  P  E  C  A  H  G  G  L  Y  K  P  V  Q  C      251
AAC GAC AAT GTG GTG ATC CCT GAG TGT GCG CAC GGC GGC CTC TAC AAG CCA GTG CAG TGC      946

H  P  S  T  G  Y  C  W  C  V  L  V  D  T  G  R  P  I  P  G      271
CAC CCC TCC ACG GGG TAC TGC TGG TGC GTC CTG GTG GAC ACG GGG CGC CCC ATT CCC GGC     1006

T  S  T  R  Y  E  Q  P  K  C  D  N  T  A  R  A  H  P  A  K      291
ACA TCC ACA AGG TAC GAG CAG CCG AAA TGT GAC AAC ACG GCC AGG GCC CAC CCA GCC AAA     1066
```

Fig. 14B

```
 A   R   D   L   Y   K   G   R   Q   L   Q   G   C   P   G   A   K   K   H   E        311
GCC CGG GAC CTG TAC AAG GGC CGC CAG CTA CAA GGT TGT CCG GGT GCC AAA AAG CAT GAG      1126

 F   L   T   S   V   L   D   A   L   S   T   D   M   V   H   A   A   S   D   P        331
TTT CTG ACC AGC GTT CTG GAC GCG CTG TCC ACG GAC ATG GTC CAC GCC GCC TCC GAC CCC      1186

 S   S   S   S   G   R   L   S   E   P   D   P   S   H   T   L   E   E   R   V        351
TCC TCC TCG TCA GGC AGG CTC TCA GAA CCC GAC CCC AGC CAT ACC CTA GAG GAG CGG GTG      1246

 V   H   W   Y   F   K   L   L   D   K   N   S   S   G   D   I   G   K   K   E        371
GTG CAC TGG TAC TTC AAA CTA CTG GAT AAA AAC TCC AGT GGA GAC ATC GGC AAA AAG GAA      1306

 I   K   P   F   K   R   F   L   R   K   K   S   K   P   K   K   C   V   K   K        391
ATC AAA CCC TTC AAG AGG TTC CTT CGC AAA AAA TCA AAG CCC AAA AAA TGT GTG AAG AAG      1366

 F   V   E   Y   C   D   V   N   N   D   K   S   I   S   V   Q   E   L   M   G        411
TTT GTT GAA TAC TGT GAC GTG AAT AAT GAC AAA TCC ATC TCC GTA CAA GAA CTG ATG GGC      1426

 C   L   G   V   A   K   E   D   G   K   A   D   T   K   K   R   H   T   P   R        431
TGC CTG GGC GTG GCG AAA GAG GAC GGC AAA GCG GAC ACC AAG AAA CGC CAC ACC CCC AGA      1486

 G   H   A   E   S   T   S   N   R   Q   P   R   K   Q   G   *                        447
GGT CAT GCT GAA AGT ACG TCT AAT AGA CAG CCA AGG AAA CAA GGA TAA                       1534

ATGGCTCATACCCCGAAGGCAGTTCCTAGACACATGGGAAATTTCCCTCACCAAAGAGCAATTAAGAAAACAAAAACAG     1613

AAACACATAGTATTTGCACTTTGTACTTTAAATGTAAATTCACTTTGTAGAAATGAGCTATTTAAACAGACTGTTTTAA     1692
```

Fig. 14C

TCTGTGAAAATGGAGAGCTGGCTTCAGAAAATTAATCACATACAATGTATGTGTCCTCTTTTGACCTTGGAAATCTGTA 1771

TGTGGTGGAGAAGTATTTGAATGCATTTAGGCTTAATTTCTTCGCCTTCCACATGTTAACAGTAGAGCTCTATGCACTC 1850

CGGCTGCAATTGTATGGCTTTCTCTAACCCCTGCAGTCACTTCCAGATGCCTGTGCTTACAGCATTGTGGAATCATGTT 1929

GGAAGCTCCACATGTCCATGGAAGTTTGTGATGTACGGCCGACCCTACAGGCAGTTAACATGCATGGGCTGGTTTGTTT 2008

CTTGGGATTTTCTGTTAGTTTGTCTTGTTTTGCTTCCAGAGATCTTGCTCATACAATGAATCACGCAACCACTAAAGC 2087

TATCCAGTTAAGTGCAGGTAGTTCCCCTGGAGGAAATAATATTTTCAAACTGTCGTTGGTGTGATACTTTGGCTCAAAG 2166

GATCTTTGCTTTTTCCATTTTAAGCTTCTGTTTTGAGTTTTGCCCTGGGGCTTGAATGAGTCCCAGAGAGTCGTTCGGAT 2245

GGTGGGAGGCTGCCTAGGAGGCAGTAAATCCAGTCACAGTGCCTGGGAGGGGCCCATCCTTCCAAAATGTAAATCCAGT 2324

CGCGGTGTGACCGAGCTGGCTAACAGGCTTGTCTGCCTGGTTTTCCTCCTACACGTGGACATTATTCTCCTGATCCTCC 2403

TACCTGGTCCACCCCAGGGCTACCGGAAGGTAAAATCTTCACCTGAACCAATTATGAGCAGTCTCCTTACTGAAGGTAC 2482

AGCCGGATACGTGGTGCCCCCGGGGCTGGTGTTGGCAGCCGGGGGGGAGGTGCCTGAGGGTCCCCACGGTTCCTTTCTGC 2561

Fig. 14D

```
TTTTCTGAATGCATCAAGGGTACGAGAACTTGCCAATGGGAAATTCATCCGAGTGGCACTGGCAGAGAAGGATAGGAGT   2640

GGAATGCCCACACAGTGACCAACAGAACTGGTCTGCGTGCATAACCAGCTGCCACCCTCAGGCCTGGGCCCCAGAGCTC   2719

AGGGCACCCAGTGTCTTAAGGAACCATTTGGAGGACAGTCTGAGAGCAGGAACTTCAAGCTGTGATTCTATCTCGGCTC   2798

AGACTTTTGGTTGGAAAAAGATCTTCATGGCCCCAAATCCCCTGAGACATGCCTTGTAGAATGATTTTGTGATGTTGTG   2877

ATGCTTGTGGAGCATCGCGTAAGGCTTCTTGCTTATTTAAACTGTGCAAGGTAAAAATCAAGCCTTTGGAGCCACAGAA   2956

CCAGCTCAAGTACATGCCAATGTTGTTTAAGAAACAGTTATGATCCTAAACTTTTTGGATAATCTTTTATATTTCTGAC   3035

CTTTGAATTTAATCATTGTTCTTAGATTAAAATAAAATATGCTATTGAAACTAAAAAAAAAAAAAAAAAAGGGCGGC   3114

CGC                                                                               3117
```

Fig. 14E

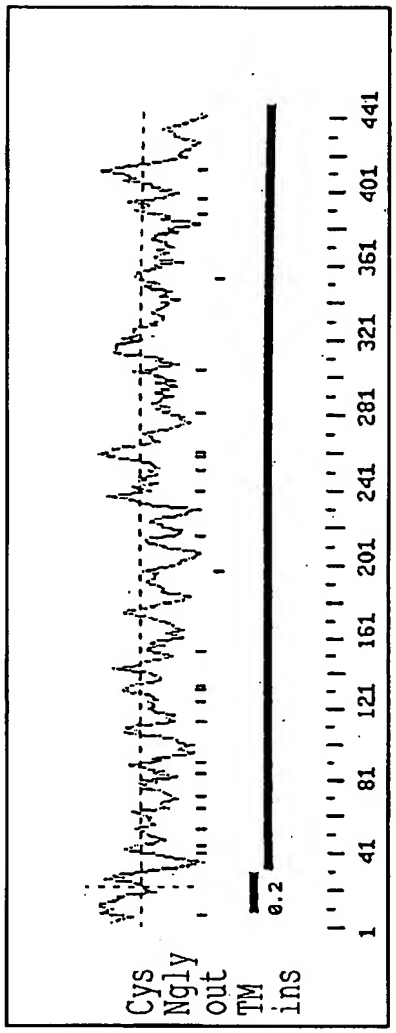

Fig. 15

```
CONCENSUS      *->YvPnCdenGfYkpvQCdpSlggqrGeCWCVDaetGkevpGtsrrarg
                 ++P+C+++G Y+ vQC+ S    +G+CWCV ++ G+++ G + +a
TANGO 209 109    FIPECNDDGTYSQVQCH-SY---TGYCWCVTPN-GRPISG-TAVAHK 149

                 dprC<-*
                   prC
               150 TPRC                                         153


CONCENSUS      *->YvPnCdenGfYkpcQCdpSlggqrGeCWCVDaetGkevpGtsrrarg
                 +P+C + G YkpvQC+pS+     G+CWCV  +tG++-pG + ++
TANGO 209 237    -IPECAHGGLYKPVQCHPST----GYCWCVLvDTGRPIPG-TSTRYE 277

                 dprC<-*
                 +p+C
               278 QPKC                                         281
```

Fig. 16

```
CONCENSUS      *->Cscypettskdgklvacpre.ydpvcgsdgvtysnecelkkaacaen
                  Cs +          C +++ +P+C+sDG+T  + Ce+++a+C
TANGO 209 40      CSLD----------CAGSpQKPLCASDGRTFLSRCEFQRAKC---  71

                  veqgtniekkhdGpC<-*
                    +  ++e++++G C
              72    --KDPQLEIAYRGNC                              84
```

Fig. 17

```
GTCGACCCACGCGTCCGGGAGCGCCGGCCAGAGCGCACGGAGGGGCCGTGCGGTCTCCACCAGCGGCCATAGGACCGCG        79

GAGCAGTTCTAGCCCGCTCGCCCGCACGTTCGCACACCGGATCTTCGCCGAGTGCCAGGGCGCAGCGCGTGGGGCGTCT       158

                                  M   L   P   P   Q   L   C   W   L   P   L   L          12
GCCTCGCTTGGTCCCCTCCAGCGTCACC ATG CTG CCG CCA CAG CTG TGC TGG CTG CCG CTG CTC            222

 A   A   L   L   P   P   V   P   A   Q   K   F   S   A   L   T   F   L   R   V         32
GCT GCC TTG CTG CCG CCG GTG CCC GCG CAG AAG TTC TCA GCG CTC ACG TTC TTG AGA GTC         282

 D   Q   D   K   D   R   D   C   S   L   D   C   P   S   S   P   Q   K   P   L         52
GAT CAA GAC AAA GAC AGA GAC TGC AGC CTG GAC TGC CCA AGC TCC CCT CAG AAG CCA CTC         342

 C   A   S   D   G   R   T   F   L   S   R   C   E   F   Q   R   A   K   C   K         72
TGT GCC TCG GAC GGG AGG ACC TTC CTG TCC CGA TGT GAG TTC CAG CGG GCC AAG TGC AAA         402

 D   P   Q   L   E   I   A   H   R   G   N   C   K   D   V   S   R   C   V   A         92
GAT CCA CAG CTG GAG ATC GCT CAC CGT GGG AAT TGC AAA GAT GTG TCC AGG TGT GTG GCT         462

 E   R   K   Y   T   Q   E   Q   A   R   K   E   F   Q   Q   V   F   I   P   E        112
GAG AGG AAG TAT ACC CAG GAG CAG GCC CGG AAG GAG TTC CAG CAA GTG TTC ATT CCA GAA         522

 C   N   D   D   G   T   Y   S   Q   V   Q   C   H   S   Y   T   G   Y   C   W        132
TGC AAT GAT GAC GGC ACC TAC AGT CAG GTC CAG TGT CAC AGC TAC ACA GGA TAC TGT TGG         582

 C   V   T   P   N   G   R   P   I   S   G   T   A   V   A   H   K   T   P   R        152
TGT GTT ACA CCA AAT GGA AGA CCC ATC AGT GGC ACT GCT GTG GCC CAC AAG ACA CCC AGG         642
```

Fig. 18A

```
 C   P   G   S   I   N   E   K   V   P   Q   R   E   G   A   G   K   A   D   D      172
TGC CCC GGT TCA ATA AAT GAA AAG GTG CCG CAG CGG GAA GGA GCA GGG AAA GCA GAT GAT      702

 A   A   A   P   A   L   E   T   Q   P   Q   G   D   E   E   D   I   A   S   R      192
GCT GCA GCC CCA GCA TTG GAG ACT CAG CCC CAA GGA GAT GAA GAA GAT ATT GCC TCA CGC      762

 Y   P   T   L   W   T   E   Q   V   K   S   R   Q   N   K   T   N   K   N   S      212
TAC CCT ACA CTC TGG ACC GAG CAA GTT AAG AGT CGG CAG AAC AAG ACC AAT AAA AAT TCA      822

 A   S   S   C   D   Q   E   H   Q   S   A   L   E   E   A   K   Q   P   K   N      232
GCA TCC TCC TGT GAT CAG GAG CAT CAG TCA GCT CTT GAG GAA GCC AAG CAG CCC AAG AAT      882

 D   N   V   V   I   P   E   C   A   H   G   G   L   Y   K   P   V   Q   C   H      252
GAC AAT GTA GTG ATC CCC GAG TGT GCA CAT GGT GGT CTC TAC AAG CCA GTG CAA TGC CAT      942

 P   S   T   G   Y   C   W   C   V   L   V   D   T   G   R   P   I   P   G   T      272
CCA TCC ACC GGA TAC TGC TGG TGT GTG CTA GTG GAC ACT GGA CGG CCC ATT CCC GGG ACC     1002

 S   T   R   Y   E   Q   P   K   C   D   N   T   A   R   A   H   P   A   K   A      292
TCC ACA AGG TAT GAG CAA CCT AAG TGT GAC AAC ACA GCC CGA GCT CAC CCA GCG AAG GCC     1062
```

Fig. 18B

```
 R   D   L   Y   K   N   R   P   L   Q   G   C   P   G   A   K   K   H   E   F        312
CGG GAC CTG TAC AAG AAC AGG CCA CTG CAG GGT TGT CCT GGT GCC AAA AAG CAC GAG TTT      1122

 L   T   S   V   L   D   A   L   S   T   D   M   V   H   A   V   S   D   P   S        332
CTG ACA AGT GTC CTG GAT GCG CTC TCC ACA GAC ATG GTT CAT GCC GTC TCT GAC CCC TCT      1182

 S   S   S   G   R   L   S   E   P   D   P   S   H   T   L   E   E   R   V   V        352
TCC TCT TCT GGC AGG CTG TCA GAG CCA GAC CCC AGC CAC ACC CTG GAG GAG AGG GTT GTA      1242

 H   W   Y   F   K   L   L   D   K   N   S   S   G   D   I   G   K   K   E   I        372
CAT TGG TAC TTC AAG CTG CTT GAT AAG AAC TCT AGT GGA GAC ATT GGC AAG AAG GAG ATC      1302

 K   P   F   K   R   F   L   R   K   K   S   K   P   K   K   C   V   K   K   F        392
AAA CCC TTT AAG AGG TTC CTG CGA AAG AAA TCC AAG CCC AAA AAG TGT GTG AAG AAG TTT      1362

 V   E   Y   C   D   M   N   N   D   K   S   I   T   V   Q   E   L   M   G   C        412
GTG GAG TAC TGC GAC ATG AAC AAT GAC AAG TCC ATC ACC GTG CAG GAG CTC ATG GGC TGC      1422

 L   G   V   T   R   E   E   G   K   A   N   T   R   K   R   H   T   P   R   G        432
TTG GGT GTC ACC AGA GAG GAG GGT AAA GCC AAC ACC AGG AAG CGC CAC ACC CCC AGA GGA      1482

 N   A   E   S   S   S   S   N   R   Q   P   R   K   Q   G   *                        448
AAT GCT GAA AGT TCT TCT TCT AAT AGA CAG CCC AGG AAA CAA GGA TGA                      1530

ACGGCTGACTACTCAAGACAGTTCCTAGACATGTGGGAATTTTCCCTCACCAAAGAGCAATTAAAAACAAAAAACATAT    1609

AGTATTTGCACTTTGTACTTTAAATGTAAATTCACTTTGTAGAAATGAGATATTTAAACGGACTGTTGTGATCTGTGAA    1688
```

Fig. 18C

AACGGAAGGCTGGCTTCGGGAAATTAATCACATACAATGTATGTGTCCTTGTTTGACCTTCAAGATCTGTGCTGGTGGG 1767

GTGGGTTTTAAATGCATTTCAACTTCACTTCCTCGTCCCTCTGTGGAGGGCCTGGTTTATTCCCTTGCCCTTTCCTCTG 1846

CCTGCAGTCTCCGTTGGCTTGCTCTCCGTGCTCTTGATGCCAGCCAGAGCCTACCGTATCCTTAGAAGCTTGCAGTGGT 1925

TGCCTGTTTCATAAAGTTTTGCTCTCTAGAGATGTCACTCACACCATGGGTCACACAGGCACTAGCTCAGGCATAGCCT 2004

TAACTTCCAGTAGCCCCTGCTTGAGACTATGCTGTTAGGCTCCAGTTTTAGCTTCTGCTTGAGCCAACCCTACTGTTAA 2083

ATCCCCGGGGAGCTGAATGGACAAGCAGCCAGAGAGTCCTCTTTCAGAGAGCACATCTAACTCTGGCCTGCCGACTCCA 2162

TCTCATTGGCCCAGGAAGGATCCAGTCTACCTGCCTCACTTCTGAGTGCCTGGGCCATCCTGCTAGGGTCCACTGGAAA 2241

GGAAAAATGCCTTCTACAGAGATCATGGACCACATCCTGAGGAAGGCACAGCTAGACTACATCGTCCCAAGAGTGCTGA 2320

ACAACAGGCAGCTGAGAATGTCAATTTCATTTTGTGGAGGCATCGGGTTTGAGGGAATTCACCCAAGGCCACTATGGGA 2399

TGAAGGTTATGAACTGGATGACCATACTGTGAGAGACTGAACTAGATTGGGTGTATGGCTGGCTGTCACTTTCTGACCT 2478

AGGTCAGTCTGTACTCTCTGTTCTCAAGGAACAGTGGCAGATGGGAGGAGAGCTGGGGCAATTCTTTCAGATTGTGGTT 2557

Fig. 18D

TATATGGGAAATTGTTCATGGTCCCCAGCCTCTCAATGTATGTCTGTAGATGCATTGTGAACGCATGCTGCAGAAGGCC 2636

TTTGCTCATTTGAACTGTGTAGAGGCAAGAAGTGACCGGCTCATGTCAATGCTGTCTACAAACGAGTATGATCCTAACT 2715

GTTTTGGATAATCTTTTATATTTCTGAACTCTGAATTTAATCATTTTATTAGATTAAAATATGCGATTAAAAAAAAAA 2794

AAAAAGGGCGGCCGCT 2810

Fig. 18E

```
          10        20        30        40        50        60        70
MURINE    ATGCTGCCGCCACAGCTGTGCTGGCTGCCGCTCGCTGCCTTGCTGCCGCCGGTGCCCGCGCAGAAGT
          :::::::  ::  :::::  :::::::::::::::::::::::  ::::  :::::::::::::::  :::::::
HUMAN     ATGCTGCTCCCCCCAGCTCTGCTGGCTGCCGCTCGCTGCCTTGCTGGGCTGCTCCCGCCGGTGCCCGCTCAGAAGT
          10        20        30        40        50        60        70

          80        90        100       110       120       130       140
TCTCAGCGCTCACGTTCTTGAGAGTCGATCAAGACAAAGACAGAGACTGCAGCCTGGACTGCCCAAGCTC
::::.:::::::::::: ::::::::: :::::::..:: :: ::: :::::::  :...: ::
TCTCGGCGCTCACGTTTTTGAGAGTGGATCAAGATAAAGACAAGGATTGTAGCTTGGACTGTGCGGGTTC
          80        90        100       110       120       130       140

          150       160       170       180       190       200       210
CCCTCAGAAGCCACTCTGTGCCTCGGACGGGAGGACCTTCCTGTCCCGATGTGAGTTCCAGCGGGCCAAG
:: ::::.::.:::: :: :: :::::.:::::::::::: :::::.:::::.:: ::.:: :::::
GCCCCAGAAACCTCTCTGCGCATCTGACGGAAGGACCTTCCTTTCCCGTTGTGAATTTCAACGTGCCAAG
          150       160       170       180       190       200       210

          220       230       240       250       260       270       280
TGCAAAGATCCACAGCTGGAGATCGCTCACCGTGGGAATTGCAAAGATGTGTCCAGGTGTGTGGCTGAGA
:::::::::: :::::.::::: ::. : ::.::.:: :::::::: :::::::::::::::::: :::.
TGCAAAGATCCCCAGCTAGAGATTGCATATCGAGGAAACTGCAAAGACGTGTCCAGGTGTGTGGCCGAAA
          220       230       240       250       260       270       280

          290       300       310       320       330       340       350
GGAAGTATACCCAGGAGCAGGCCCCGGAAGGAGTTCCAGCAAGTGTTCATTCCAGAATGCAATGATGACGG
:::::::::::::::::::.:::::::::::::::: ::::::::::::::::::..::.:::::::: :::::
GGAAGTATACCCAGGAGCAAGCCCCGGAAGGAGTTTCAGCAAGTGTTCATTCCTGAGTGCAATGACGACGG
          290       300       310       320       330       340       350
```

Fig. 19A

```
      360       370       380       390       400       410       420
CACCTACAGTCAGGTCCAGTGTCACAGCTACACAGGATACTGTTGGTGTGTTACACCAAATGGAAGACCC
:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
CACCTACAGTCAGGTCCAGTGTCACAGCTACACGGGATACTGCTGGTGCGTCACGCCCAACGGGAGGCCC
      360       370       380       390       400       410       420

      430       440       450       460       470       480       490
ATCAGTGGCACTGCTGTGGCCCACAAGACACCCAGGTGCCCCGGTTCAATAAATGAAAAGGTGCCGCAGC
::::: :::::::: :::::::::::::::::.::: ::::::: ::::: .:::::::::::: :.::: ::.:
ATCAGCGGCACTGCCGTGGCCCACAAGACGCCCCGGTGCCCGGGTTCCGTAAATGAAAAGTTACCCCAAC
      430       440       450       460       470       480       490

      500       510       520       530       540       550       560
GGGAAGGAGCAGGGAAAGCAGATGATGCTGCAGCCCCCAGCATTGGAGACTCAGCCCCCAAGGAGATGAAGA
: ::::: .::::.:::.:::::::::::: ::::: :::::.:::::::::::::::: :::::::::::::
GCGAAGGCACAGGAAAAACAGATGATGCCGCAGCTCCAGCGTTGGAGACTCAGCCTCAAGGAGATGAAGA
      500       510       520       530       540       550       560

      570       580       590       600       610       620       630
AGATATTGCCTCACGCTACCCTACACTCTGGACCGAGCAAGTTAAGAGTCGGCAGAACAAGACCAATAAA
::::::::: ::::: :::::::: :: ::::: ::.::.:::::.:::::::::::::::.:::::::.
AGATATTGCATCACGTTACCCTACCCTTTGGACTGAACAGGTTAAAAGTCGGCAGAACAAAACCAATAAG
      570       580       590       600       610       620       630

      640       650       660       670       680       690       700
AATTCAGCATCCTCCTGTGATCAGGAGCATCAGTCAGCTCTTGAGGAAGCCAAGCAGCCCAAGAATGACA
::::::: .:: :::::::::.::.:::: :::::.:: :: ::::::::::::::::::::::: ::::
AATTCAGTGTCATCCTGTGACCAAGAGCACCAGTCTGCCCTGGAGGAAGCCAAGCAGCCCAAGAACGACA
      640       650       660       670       680       690       700
```

Fig. 19B

```
      710       720       730       740       750       760       770
ATGTAGTGATCCCCGAGTGTGCACATGGTGGTCTCTACAAGCCAGTGCAATGCCATCCATCCACCGGATA
:::::::::::: ::::::::::.:: :: :: :::::::::::::::::::.::::: :: ::::: ::.::
ATGTGGTGATCCCTGAGTGTGCGCACGGCGGCCTCTACAAGCCAGTGCAGTGCCACCCCTCCACGGGGTA
      710       720       730       740       750       760       770

      780       790       800       810       820       830       840
CTGCTGGTGTGTGCTAGTGGACACTGGACGGCCCATTCCCGGGACCTCCACAAGGTATGAGCAACCTAAG
::::::::: :: ::.::::::::: ::.:: :::::::::::: :: :::::::::: :::::.:: ::.
CTGCTGGTGCGTCCTGGTGGACACGGGGCGCCCCATTCCCGGCACATCCACAAGGTACGAGCAGCCGAAA
      780       790       800       810       820       830       840

      850       860       870       880       890       900       910
TGTGACAACACAGCCCGAGCTCACCCAGCGAAGGCCCGGGACCTGTACAAGAACAGGCCACTGCAGGGTT
:::::::::::::.::: :.:: :::::::: ::.::::::::::::::::::..: : :.:::.::.:::
TGTGACAACACGGCCCAGGGCCCACCCAGCCAAAGCCCGGGACCTGTACAAGGGCCGCCAGCTACAAGGTT
      850       860       870       880       890       900       910

      920       930       940       950       960       970       980
GTCCTGGTGCCAAAAAGCACGAGTTTCTGACAAGTGTCCTGGATGCGCTCTCCACAGACATGGTTCATGC
:::: :::::::::::::: :::::::::: :: :: ::::: :::::.:::::::: :: ::
GTCCGGGTGCCAAAAAGCATGAGTTTCTGACCAGCGCTTCTGGACGCGCTGTCCACGGACATGGTCCACGC
      920       930       940       950       960       970       980

      990      1000      1010      1020      1030      1040      1050
CGTCTCTGACCCCTCTTCCTCTTCTGGCAGGCTGTCAGAGCCAGACCCCAGCCACACCCTGGAGGAGAGG
:: ::: :::::::: ::::: ::.::::::::: :::::.:: :::::::::::: :::::.:::::: ::
CGCCTCCGACCCCTCCTCCTCGTCAGGCAGGCTCTCAGAACCCGACCCCAGCCATACCCTAGAGGAGCGG
      990      1000      1010      1020      1030      1040      1050
```

Fig. 19C

```
      1060       1070       1080       1090       1100       1110       1120
GTTGTACATTGGTACTTCAAGCTGCTTGATAAGAACTCTAGTGGAGACATTGGCAAGAAGGAGATCAAAC
:: ::.:: ::::::::::::.::.:: :::::.::::: :::::::::: :::::.:::::.:::::::
GTGGTGCACTGGTACTTCAAACTACTGGATAAAAACTCCAGTGGAGACATCGGCAAAAAGGAAATCAAAC
      1060       1070       1080       1090       1100       1110       1120

      1130       1140       1150       1160       1170       1180       1190
CCTTTAAGAGGTTCCTGCGAAAGAAATCCAAGCCCAAAAAGTGTGTGAAGAAGTTTGTGGAGTACTGCGA
:::: :::::::::: :: ::.::::: ::::::::::.:::::::::::::::::: ::.::::: ::
CCTTCAAGAGGTTCCTTCGCAAAAAATCAAAGCCCAAAAAATGTGTGAAGAAGTTTGTTGAATACTGTGA
      1130       1140       1150       1160       1170       1180       1190

      1200       1210       1220       1230       1240       1250       1260
CATGAACAATGACAAGTCCATCACCGTGCAGGAGCTCATGGGCTGCTTGGGTGTCACCAGAGAGGAGGGT
:.:::: ::::::::.::::::.::::.::.::.:: ::::::::: :::: :: .: :.::::::: ::
CGTGAATAATGACAAATCCATCTCCGTACAAGAACTGATGGGCTGCCTGGGCGTGGCGAAAGAGGACGGC
      1200       1210       1220       1230       1240       1250       1260

      1270       1280       1290       1300       1310       1320       1330
AAAGCCAACACCAGGAAGCGCCACACCCCCAGAGGAAATGCTGAAAGTTCTTCTTCTAATAGACAGCCCA
::::: .:::::::.:::.:::::::::::::::::. ::::::::::::.: :::   :::::::::::: :
AAAGCGGACACCAAGAAACGCCACACCCCCAGAGGTCATGCTGAAAGTACGTCT---AATAGACAGCCAA
      1270       1280       1290       1300       1310          1320

      1340
GGAAACAAGGA
::::::::::::
GGAAACAAGGA
1330
```

Fig. 19D

```
        10        20        30        40        50        60        70
MLLPQLCWLPLLAGLLPPVPAQKFSALTFLRVDQDKDKDCSLDCAGSPQKPLCASDGRTFLSRCEFQRAK
:: ::::::::::.:::::::::::::::::::::.:::::..:::::::::::::::::::::::::::
MLPPQLCWLPLLAALLPPVPAQKFSALTFLRVDQDKDRDCSLDCPSSPQKPLCASDGRTFLSRCEFQRAK
        10        20        30        40        50        60        70

        80        90       100       110       120       130       140
CKDPQLEIAYRGNCKDVSRCVAERKYTQEQARKEFQQVFIPECNDDGTYSQVQCHSYTGYCWCVTPNGRP
::::::::: ::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
CKDPQLEIAHRGNCKDVSRCVAERKYTQEQARKEFQQVFIPECNDDGTYSQVQCHSYTGYCWCVTPNGRP
        80        90       100       110       120       130       140

       150       160       170       180       190       200       210
ISGTAVAHKTPRCPGSVNEKLPQREGTGKTDDAAAPALETQPQGDEEDIASRYPTLWTEQVKSRQNKTNK
:::::::::::::::::.:::.:::::.::.:::::::::::::::::::::::::::::::::::::::
ISGTAVAHKTPRCPGSINEKVPQREGAGKADDAAAPALETQPQGDEEDIASRYPTLWTEQVKSRQNKTNK
       150       160       170       180       190       200       210

       220       230       240       250       260       270       280
NSVSSCDQEHQSALEEAKQPKNDNVVIPECAHGGLYKPVQCHPSTGYCWCVLVDTGRPIPGTSTRYEQPK
::.:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
NSASSCDQEHQSALEEAKQPKNDNVVIPECAHGGLYKPVQCHPSTGYCWCVLVDTGRPIPGTSTRYEQPK
       220       230       240       250       260       270       280

       290       300       310       320       330       340       350
CDNTARAHPAKARDLYKGRQLQGCPGAKKHEFLTSVLDALSTDMVHAASDPSSSSGRLSEPDPSHTLEER
:::::::::::::::::.:.:::::::::::::::::::::::::::.:::::::::::::::::::::::
CDNTARAHPAKARDLYKNRPLQGCPGAKKHEFLTSVLDALSTDMVHAVSDPSSSSGRLSEPDPSHTLEER
       290       300       310       320       330       340       350
```

Fig. 20A

```
      360       370       380       390       400       410       420
VVHWYFKLLDKNSSGDIGKKEIKPFKRFLRKKSKPKKCVKKFVEYCDVNNDKSISVQELMGCLGVAKEDG
:::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
VVHWYFKLLDKNSSGDIGKKEIKPFKRFLRKKSKPKKCVKKFVEYCDMNNDKSITVQELMGCLGVTREEG
      360       370       380       390       400       410       420

      430       440
KADTKKRHTPRGHAESTS-NRQPRKQG
::.:.:::::::::.:::.: :::::::
KANTRKRHTPRGNAESSSSNRQPRKQG
      430       440
```

Fig. 20B

```
CGCTTACTCCTTTGCCTTCGCAAACAGGGAAAAGTGTTCCACGAAGCGGTAGCGCCTTTCCGCCTCGCGTTTTCCTCCC      79

TGACCCTGGTCCCGGCTCCCGTCCGGGCGCCAGCTGGTGGGGCGAGCGCCGGGAGCCCATCTGCCCCCAGGGGCACGGG     158

GCGCGGGGCCGGCTCCCGCCCGGCACATGGCTGCAGCCACCTCGCGCGCACCCCGAGGCGCCGCGCCCAGCTCGCCCGA     237

GGTCCGTCGGAGGCGCCCGGCCGCCCCGGAGCCAAGCAGCAACTGAGCGGGGAAGCGCCCGCGTCCGGGGATCGGG      313

 M   S   L   L   L   L   L   L   L   V   S   Y   Y   V   G   T   L   G   T   H      20
ATG TCC CTC CTC CTT CTC CTC TTG CTA GTT TCC TAC TAT GTT GGA ACC TTG GGG ACT CAC      373

 T   E   I   K   R   V   A   E   E   K   V   T   L   P   C   H   H   Q   L   G      40
ACT GAG ATC AAG AGA GTG GCA GAG GAA AAG GTC ACT TTG CCC TGC CAC CAT CAA CTG GGG      433

 L   P   E   K   D   T   L   D   I   E   W   L   L   T   D   N   E   G   N   Q      60
CTT CCA GAA AAA GAC ACT CTG GAT ATT GAA TGG CTG CTC ACC GAT AAT GAA GGG AAC CAA      493

 K   V   V   I   T   Y   S   S   R   H   V   Y   N   N   L   T   E   E   Q   K      80
AAA GTG GTG ATC ACT TAC TCC AGT CGT CAT GTC TAC AAT AAC TTG ACT GAG GAA CAG AAG      553

 G   R   V   A   F   A   S   N   F   L   A   G   D   A   S   L   Q   I   E   P     100
GGC CGA GTG GCC TTT GCT TCC AAT TTC CTG GCA GGA GAT GCC TCC TTG CAG ATT GAA CCT      613

 L   K   P   S   D   E   G   R   Y   T   C   K   V   K   N   S   G   R   Y   V     120
CTG AAG CCC AGT GAT GAG GGC CGG TAC ACC TGT AAG GTT AAG AAT TCA GGG CGC TAC GTG      673

 W   S   H   V   I   L   K   V   L   V   R   P   S   K   P   K   C   E   L   E     140
TGG AGC CAT GTC ATC TTA AAA GTC TTA GTG AGA CCA TCC AAG CCC AAG TGT GAG TTG GAA      733

 G   E   L   T   E   G   S   D   L   T   L   Q   C   E   S   S   S   G   T   E     160
GGA GAG CTG ACA GAA GGA AGT GAC CTG ACT TTG CAG TGT GAG TCA TCC TCT GGC ACA GAG      793
```

Fig. 21A

```
  P   I   V   Y   Y   W   Q   R   I   R   E   K   E   G   E   D   E   R   L   P      180
CCC ATT GTG TAT TAC TGG CAG CGA ATC CGA GAG AAA GAG GGA GAG GAT GAA CGT CTG CCT      853

  P   K   S   R   I   D   Y   N   H   P   G   R   V   L   L   Q   N   L   T   M      200
CCC AAA TCT AGG ATT GAC TAC AAC CAC CCT GGA CGA GTT CTG CTG CAG AAT CTT ACC ATG      913

  S   Y   S   G   L   Y   Q   C   T   A   G   N   E   A   G   K   E   S   C   V      220
TCC TAC TCT GGA CTG TAC CAG TGC ACA GCA GGC AAC GAA GCT GGG AAG GAA AGC TGT GTG      973

  V   R   V   T   V   Q   Y   V   Q   S   I   G   M   V   A   G   A   V   T   G      240
GTG CGA GTA ACT GTA CAG TAT GTA CAA AGC ATC GGC ATG GTT GCA GGA GCA GTG ACA GGC     1033

  I   V   A   G   A   L   L   I   F   L   L   V   W   L   L   I   R   R   K   D      260
ATA GTG GCT GGA GCC CTG CTG ATT TTC CTC TTG GTG TGG CTG CTA ATC CGA AGG AAA GAC     1093

  K   E   R   Y   E   E   E   E   R   P   N   E   I   R   E   D   A   E   A   P      280
AAA GAA AGA TAT GAG GAA GAA GAG AGA CCT AAT GAA ATT CGA GAA GAT GCT GAA GCT CCA     1153
```

Fig. 21B

```
 K   A   R   L   V   K   P   S   S   S   S   S   G   S   R   S   S.  R   S   G        300
AAA GCC CGT CTT GTG AAA CCC AGC TCC TCT TCC TCA GGC TCT CGG AGC TCA CGC TCT GGT      1213

 S   S   S   T   R   S   T   A   N   S   A   S   R   S   Q   R   T   L   S   T        320
TCT TCC TCC ACT CGC TCC ACA GCA AAT AGT GCC TCA CGC AGC CAG CGG ACA CTG TCA ACT      1273

 D   A   A   P   Q   P   G   L   A   T   Q   A   Y   S   L   V   G   P   E   V        340
GAC GCA GCA CCC CAG CCA GGG CTG GCC ACC CAG GCA TAC AGC CTA GTG GGG CCA GAG GTG      1333

 R   G   S   E   P   K   K   V   H   H   A   N   L   T   K   A   E   T   T   P        360
AGA GGT TCT GAA CCA AAG AAA GTC CAC CAT GCT AAT CTG ACC AAA GCA GAA ACC ACA CCC      1393

 S   M   I   P   S   Q   S   R   A   F   Q   T   V   *                              374
AGC ATG ATC CCC AGC CAG AGC AGA GCC TTC CAA ACG GTC TGA                              1435

ATTACAATGGACTTGACTCCCACGCTTTCCTAGGAGTCAGGGTCTTTGGACTCTTCTCGTCATTGGAGCTCAAGTCACC    1514

AGCCACACAACCAGATGAGAGGTCATCTAAGTAGCAGTGAGCATTGCACGGAACAGATTCAGATGAGCATTTTCCTTAT    1593

ACAATACCAAACAAGCAAAAGGATGTAAGCTGATTCATCTGTAAAAAGGCATCTTATTGTGCCTTTAGACCAGAGTAAG    1672

GGAAAGCAGGAGTCCAAATCTATTTGTTGACCAGGACCTGTGGTGAAGAAAGGTTGGGGAAAGGTGAGGTGAATATACC    1751

TAAAACTTTTAATGTGGGATATTTTGTATCAGTGCTTTGATTCACAATTTTCAAGAGGAAATGGGATGCTGTTTGTAAA    1830

TTTTCTATGCATTTCTGCAAACTTATTGGATTATTAGTTATTCAGACAGTCAAGCAGAACCCNCAGCCTTATTACNCCT    1909

GTCTACACCATGTACTGAGCTAACCACTTTTAAGAAACT                                              1948
```

Fig. 21C

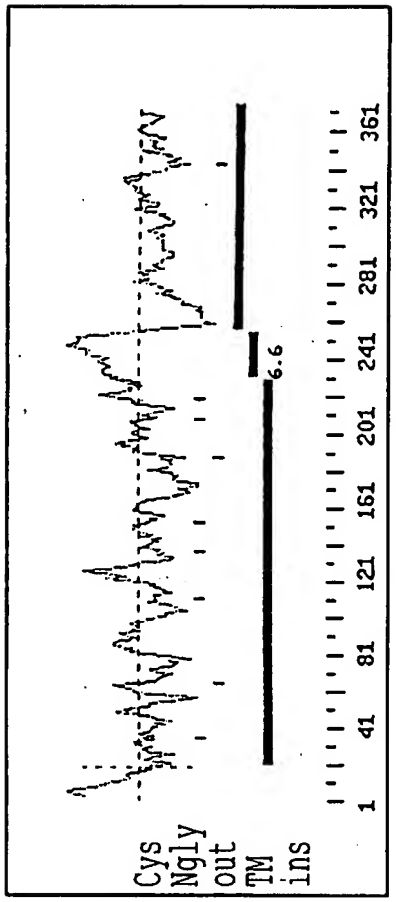

Fig. 22

```
CONCENSUS     *->GesvtLtCsvs..gfgppgvsvtWyfkngk.lgpsllgysysrlesg
                e+vtL+C     +    +  ++ W+   + + ++++++++++sr++ +
A236      28     EEKVTLPCHHQlgLPEKDTLDIEWL-LTDNeGNQKVVITYSSRHVYN  73

                ekanlsegrfsis.....sltLtissvekeDsGtYtCvv<-*
                    ++gr+ + ++   - ++L+i +++++D+G YtC v
              74 NLTEEQKGRVAFAsnflagDASLQIEPLKPSDEGRYTCKV       113


CONCENSUS     *->GesvtLtCsvsgfgppgvsvtWyf.kngk.lgpsllgysysrlesge
                G++++tL+C++s +g+ ++ ++W + ++  ++          +rl ++
A236     146    GSDLTLQCESS-SGTEPIVYYWQRiREKEgED---------ERLPPK- 182

                kanlsegrfsis.sltLtissvekeDsGtYtCvv<-*
                    r  +++  ++++++++++ sG Y C++
             183 ------SRIDYNhPGRVLLQNLTMSYSGLYQCTA              210
```

Fig. 23

```
gtcgacccac gcgtccggtt ccacgaagcg gtagctcctt gccgcctcgc cttctcctcc   60
ctaaccctgg gcccggcccc cgtcccggcg cgagctggtg gagccagggc tagaagccct  120
cggtgccccc ggagcgcagc gcgcaggggga cccgggcgcg gggccagcgc ccgcacatgg  180
ctgcagcccc ccgcgcgcac cccgaggcgc cgcgccctgc tcacagaagg tccgtcggct  240
gggctcggtc gccctgcagc caggctgcgc tgagccggga agtgcccgtg tccggagatc  300
ggg atg tcc ctc ttc ctc tgg cta gta tcc tat tat gtt gga acg          348
    Met Ser Leu Phe Phe Leu Trp Leu Val Ser Tyr Tyr Val Gly Thr
     1               5                  10                  15

ctg gga act cac act gag atc aag aga gtg gca gag gaa aag gtt acc     396
Leu Gly Thr His Thr Glu Ile Lys Arg Val Ala Glu Glu Lys Val Thr
                20                  25                  30

ttg ccc tgt cac cat caa ctg ggg ctt ccc gag aaa gac acc ctg gac     444
Leu Pro Cys His His Gln Leu Gly Leu Pro Glu Lys Asp Thr Leu Asp
            35                  40                  45

att gaa tgg ctg ctc acc gat aat gaa ggg aac caa aaa gtg gtt att     492
Ile Glu Trp Leu Leu Thr Asp Asn Glu Gly Asn Gln Lys Val Val Ile
        50                  55                  60

acg tat tcc agc cgt cat gtc tac aat aac ttg acc gag gag cag aag     540
Thr Tyr Ser Ser Arg His Val Tyr Asn Asn Leu Thr Glu Glu Gln Lys
    65                  70                  75

ggc cga gtg gcc ttc gct tcc aac ttc ctg gca gga gat gct tcc ctg     588
Gly Arg Val Ala Phe Ala Ser Asn Phe Leu Ala Gly Asp Ala Ser Leu
80                  85                  90                  95
```

Fig. 24A

```
cag att gag cct ctg aaa ccc agt gat gaa ggc aga tac acc tgc aag   636
Gln Ile Glu Pro Leu Lys Pro Ser Asp Glu Gly Arg Tyr Thr Cys Lys
            100                 105                 110

gtg aag aat tca gga cgc tat gtc tgg agc cat gtc atc ttg aaa gtg   684
Val Lys Asn Ser Gly Arg Tyr Val Trp Ser His Val Ile Leu Lys Val
        115                 120                 125

cta gtg aga cca tcc aag ccc aag tgt gag ctg gaa gga gag ccg acc   732
Leu Val Arg Pro Ser Lys Pro Lys Cys Glu Leu Glu Gly Glu Pro Thr
    130                 135                 140

gaa gga agt gac ctg acg ctg cag tgt gag tct gcc tct gga act aag   780
Glu Gly Ser Asp Leu Thr Leu Gln Cys Glu Ser Ala Ser Gly Thr Lys
145                 150                 155

ccc att gtg tat tat tgg cag cga atc cgg gag aag gag gga gaa gat   828
Pro Ile Val Tyr Tyr Trp Gln Arg Ile Arg Glu Lys Glu Gly Glu Asp
160                 165                 170                 175

gaa cac ctg cca ccc aaa tcc aga att gat tac aac aac cct ggc cga   876
Glu His Leu Pro Pro Lys Ser Arg Ile Asp Tyr Asn Asn Pro Gly Arg
            180                 185                 190

gtg ctg ctg cag aat ctc acc atg gcc tcc tct ggg ctt tac cag tgc   924
Val Leu Leu Gln Asn Leu Thr Met Ala Ser Ser Gly Leu Tyr Gln Cys
        195                 200                 205

aca gca ggc aac gag gct gga aag gag agc tgt gtg gta cgg gtg act   972
Thr Ala Gly Asn Glu Ala Gly Lys Glu Ser Cys Val Val Arg Val Thr
    210                 215                 220
```

Fig. 24B

```
gta cag tat gtg cag agc att ggc atg gtg gca gga gca gtg aca ggc      1020
Val Gln Tyr Val Gln Ser Ile Gly Met Val Ala Gly Ala Val Thr Gly
    225                 230                 235

ata gtg gca gga gcc ctg ctc att ttc ctc ctg ata tgg ctg cta ata      1068
Ile Val Ala Gly Ala Leu Leu Ile Phe Leu Leu Ile Trp Leu Leu Ile
240                 245                 250                 255

cga agg aaa agc aaa gac aga tac gag gaa gaa gac aga cct aat gaa      1116
Arg Arg Lys Ser Lys Asp Arg Tyr Glu Glu Glu Asp Arg Pro Asn Glu
                260                 265                 270

atc cga gaa gac gcc gaa gcg ccc cga gcc cgc ctt gtg aag cct agc      1164
Ile Arg Glu Asp Ala Glu Ala Pro Arg Ala Arg Leu Val Lys Pro Ser
            275                 280                 285

tcc tct tcc tca ggc tcc cgg agc tca cgc tct ggc tcc tcc tcc acc      1212
Ser Ser Ser Ser Gly Ser Arg Ser Ser Arg Ser Gly Ser Ser Ser Thr
        290                 295                 300

cgc tcc acc ggg aac agt gcc tcc aga agc cag cgg acg ctg tcg agt      1260
Arg Ser Thr Gly Asn Ser Ala Ser Arg Ser Gln Arg Thr Leu Ser Ser
    305                 310                 315

gaa gca gcg ccg cag cag ccc ggg cta gcc ccg cag gca tac agc ctc      1308
Glu Ala Ala Pro Gln Gln Pro Gly Leu Ala Pro Gln Ala Tyr Ser Leu
320                 325                 330                 335
```

Fig. 24C

```
ata gga ccg gaa gtg aga ggt tct gaa cca aag aaa gtc cac cat acg    1356
Ile Gly Pro Glu Val Arg Gly Ser Glu Pro Lys Lys Val His His Thr
            340                 345                 350

acc ctg acc aaa gca gaa acc aca ctc agc aca acg ccc agc cag agc    1404
Thr Leu Thr Lys Ala Glu Thr Thr Leu Ser Thr Thr Pro Ser Gln Ser
        355                 360                 365

aaa gcc ttc caa act gtc tga ct tagagtggac ttgacttgcg cttgccccaa    1457
Lys Ala Phe Gln Thr Val *
    370

agtcaggatc ttagcctagt cactggagct cgtccaccag ccacgcaagc ccctcagcca    1517
gatacgatct cacttaagta gctgcagaaa tggcacggac cagttctgat gagtaccctc    1577
cttatatagg ataccaaaca aacacaagga cggaggctga ccatctatct ctaaaggcac    1637
ctcactgtgc cttcagacag agtggagggg aggaggggcc caagcttatt tggtgaaaat    1697
aaagggaaag gtgaggctgc acacacctga aacatcttac ctaggatgtt gcaagtcacc    1757
acagtcaaga agaagcggga atctcgtaga tcaatttctt attcatttct gcaaatttat    1817
tggattagtg tgattattca gatagtcaaa acagaagccc acgccttata atatacctat    1877
ctgcaacatg tactgggaga actgcgttta agaaattcac attaaaaaaa aaaaaaaaaa    1937
aagggcggcc gc                                                       1949
```

Fig. 24D

```
              .         .         .         .         .
  1 ATGTCCCTCTTC...TTCCTCTGGCTAGTATCCTATTATGTTGGAACGCT  47
    ||||||||| ||    |||||| |||||| ||||| |||||||||||  |
  1 ATGTCCCTCCTCCTTCTCCTCTTGCTAGTTTCCTACTATGTTGGAACCTT  50

              .         .         .         .         .
 48 GGGAACTCACACTGAGATCAAGAGAGTGGCAGAGGAAAAGGTTACCTTGC  97
    ||| |||||||||||||||||||||||||||||||||||||| || ||||
 51 GGGGACTCACACTGAGATCAAGAGAGTGGCAGAGGAAAAGGTCACTTTGC 100

              .         .         .         .         .
 98 CCTGTCACCATCAACTGGGGCTTCCCGAGAAAGACACCCTGGACATTGAA 147
    |||| |||||||||||||||||||| || |||||||| ||||| ||||||
101 CCTGCCACCATCAACTGGGGCTTCCAGAAAAAGACACTCTGGATATTGAA 150

              .         .         .         .         .
148 TGGCTGCTCACCGATAATGAAGGGAACCAAAAAGTGGTTATTACGTATTC 197
    |||||||||||||||||||||||||||||||||||||| || || || ||
151 TGGCTGCTCACCGATAATGAAGGGAACCAAAAAGTGGTGATCACTTACTC 200

              .         .         .         .         .
198 CAGCCGTCATGTCTACAATAACTTGACCGAGGAGCAGAAGGGCCGAGTGG 247
    ||| |||||||||||||||||||||| |||||| ||||||||||||||||
201 CAGTCGTCATGTCTACAATAACTTGACTGAGGAACAGAAGGGCCGAGTGG 250

              .         .         .         .         .
248 CCTTCGCTTCCAACTTCCTGGCAGGAGATGCTTCCCTGCAGATTGAGCCT 297
    |||| |||||||| ||||||||||||||||| ||| |||||||||| |||
251 CCTTTGCTTCCAATTTCCTGGCAGGAGATGCCTCCTTGCAGATTGAACCT 300

              .         .         .         .         .
298 CTGAAACCCAGTGATGAAGGCAGATACACCTGCAAGGTGAAGAATTCAGG 347
    ||||| ||||||||||| ||| | |||||||| ||||| |||||||||||
301 CTGAAGCCCAGTGATGAGGGCCGGTACACCTGTAAGGTTAAGAATTCAGG 350
```

Fig. 25A

```
348 ACGCTATGTCTGGAGCCATGTCATCTTGAAAGTGCTAGTGAGACCATCCA 397
    |||||  ||  ||||||||||||||||||  ||||||  ||||||||||||||||||
351 GCGCTACGTGTGGAGCCATGTCATCTTAAAAGTCTTAGTGAGACCATCCA 400

398 AGCCCAAGTGTGAGCTGGAAGGAGAGCCGACCGAAGGAAGTGACCTGACG 447
    |||||||||||||||||  |||||||||||||||  |||  ||||||||||||||||||||
401 AGCCCAAGTGTGAGTTGGAAGGAGAGCTGACAGAAGGAAGTGACCTGACT 450

448 CTGCAGTGTGAGTCTGCCTCTGGAACTAAGCCCATTGTGTATTATTGGCA 497
    ||||||||||||||  ||||||||  ||  ||||||||||||||||||  ||||||
451 TTGCAGTGTGAGTCATCCTCTGGCACAGAGCCCATTGTGTATTACTGGCA 500

498 GCGAATCCGGGAGAAGGAGGGAGAAGATGAACACCTGCCACCCAAATCCA 547
    ||||||||||  ||||||  ||||||||  ||||||  ||||||||||  |
501 GCGAATCCGAGAGAAAGAGGGAGAGGATGAACGTCTGCCTCCCAAATCTA 550

548 GAATTGATTACAACAACCCTGGCCGAGTGCTGCTGCAGAATCTCACCATG 597
    |  ||||||  |||||||  ||||||  ||||||||||||||||  |||||||
551 GGATTGACTACAACCACCCTGGACGAGTTCTGCTGCAGAATCTTACCATG 600

598 GCCTCCTCTGGGCTTTACCAGTGCACAGCAGGCAACGAGGCTGGAAAGGA 647
    |||  ||||||  ||  ||||||||||||||||||||||  ||||||  ||||||
601 TCCTACTCTGGACTGTACCAGTGCACAGGCAACGAAGCTGGGAAGGA 650

648 GAGCTGTGTGGTACGGGTGACTGTACAGTATGTGCAGAGCATTGGCATGG 697
    ||||||||||||  ||  ||  ||||||||||||||  ||  ||||||  |||||||
651 AAGCTGTGTGGTGCGAGTAACTGTACAGTATGTACAAAGCATCGGCATGG 700
```

Fig. 25B

```
698 TGGCAGGAGCAGTGACAGGCATAGTGGCAGGAGCCCTGCTCATTTCCTC 747
    | ||||||||||||||||||||||||||||| |||||||||||| |||||||||||
701 TTGCAGGAGCAGTGACAGGCATAGTGGCTGGAGCCCTGCTGATTTTCCTC 750

748 CTGATATGGCTGCTAATACGAAGGAAAAGCAAAGACAGATACGAGGAAGA 797
      || | |||||||||||| |||||||| |||||| |||||||||
751 TTGGTGTGGCTGCTAATCCGAAGGAAAGACAAAGAAAGATATGAGGAAGA 800

798 AGACAGACCTAATGAAATCCGAGAAGACGCCGAAGCGCCCCGAGCCCGCC 847
    ||| |||||||||||||| |||||||| || |||||  ||   |||||| |
801 AGAGAGACCTAATGAAATTCGAGAAGATGCTGAAGCTCCAAAAGCCCGTC 850

848 TTGTGAAGCCTAGCTCCTCTTCCTCAGGCTCCCGGAGCTCACGCTCTGGC 897
    ||||||| || |||||||||||||||||| |||||||||||||||||||
851 TTGTGAAACCCAGCTCCTCTTCCTCAGGCTCTCGGAGCTCACGCTCTGGT 900

898 TCCTCCTCCACCCGCTCCACCGGGAACAGTGCCTCCAGAAGCCAGCGGAC 947
    || |||||||| |||||||||| |  || |||||||||  | ||||||||||
901 TCTTCCTCCACTCGCTCCACAGCAAATAGTGCCTCACGCAGCCAGCGGAC 950

948 GCTGTCGAGTGAAGCAGCGCCGCAGCAGCCCGGGCTAGCCCCCGCAGGCAT 997
      |||||.| ||| ||||| ||     ||||| |||||| ||| | |||||||
951 ACTGTCAACTGACGGCAGCACC...CCAGCCAGGGCTGGCCACCCCAGGCAT 997

998 ACAGCCTCATAGGACCGGAAGTGAGAGGTTCTGAACCAAAGAAAGTCCAC 1047
    |||||||  |  ||  ||  |||||||||||||||||||||||||||||
998 ACAGCCTAGTGGGGCCAGAGGTGAGAGGTTCTGAACCAAAGAAAGTCCAC 1047
```

Fig. 25C

```
1048 CATACGACCCTGACCAAAGCAGAAACCACACTCAGCACAACGCCCAGCCA 1097
     ||| | | | |||||||||||||||||||| |||||  |  |||||||||
1048 CATGCTAATCTGACCAAAGCAGAAACCACACCCAGCATGATCCCCAGCCA 1097

1098 GAGCAAAGCCTTCCAAACTGTC 1119
     ||||| ||||||||||||| |||
1098 GAGCAGAGCCTTCCAAACGGTC 1119
```

Fig. 25D

```
  1 .MSLFFLWLVSYYVGTLGTHTEIKRVAEEKVTLPCHHQLGLPEKDTLDIE 49
    |  |  ||||||||||||||||||||||||||||||||||||||||||||
  1 MSLLLLLLVSYYVGTLGTHTEIKRVAEEKVTLPCHHQLGLPEKDTLDIE 50

 50 WLLTDNEGNQKVVITYSSRHVYNNLTEEQKGRVAFASNFLAGDASLQIEP 99
    ||||||||||||||||||||||||||||||||||||||||||||||||||
 51 WLLTDNEGNQKVVITYSSRHVYNNLTEEQKGRVAFASNFLAGDASLQIEP 100

100 LKPSDEGRYTCKVKNSGRYVWSHVILKVLVRPSKPKCELEGEPTEGSDLT 149
    |||||||||||||||||||||||||||||||||||||||||||||||||
101 LKPSDEGRYTCKVKNSGRYVWSHVILKVLVRPSKPKCELEGELTEGSDLT 150

150 LQCESASGTKPIVYYWQRIREKEGEDEHLPPKSRIDYNNPGRVLLQNLTM 199
    |||||.|||.|||||||||||||||||| ||||||||||.|||||||||
151 LQCESSSGTEPIVYYWQRIREKEGEDERLPPKSRIDYNHPGRVLLQNLTM 200

200 ASSGLYQCTAGNEAGKESCVVRVTVQYVQSIGMVAGAVTGIVAGALLIFL 249
    . ||||||||||||||||||||||||||||||||||||||||||||||||
201 SYSGLYQCTAGNEAGKESCVVRVTVQYVQSIGMVAGAVTGIVAGALLIFL 250

250 LIWLLIRRKSKDRYEEEDRPNEIREDAEAPRARLVKPSSSSSGSRSSRSG 299
    |:|||||||| |:||||||:|||||||||||||:|||||||||||||||||
251 LVWLLIRRKDKERYEEEERPNEIREDAEAPKARLVKPSSSSSGSRSSRSG 300

300 SSSTRSTGNSASRSQRTLSSEAAPQQPGLAPQAYSLIGPEVRGSEPKKVH 349
    |||||||| ||||||||||||.:||| ||||| ||||||:|||||||||||
301 SSSTRSTANSASRSQRTLSTDAAP.QPGLATQAYSLVGPEVRGSEPKKVH 349
```

Fig. 26A

```
350 HTTLTKAETTLSTTPSQSKAFQTV 373
    |  |||||||  |  ||||:||||||
350 HANLTKAETTPSMIPSQSRAFQTV 373
```

Fig. 25B

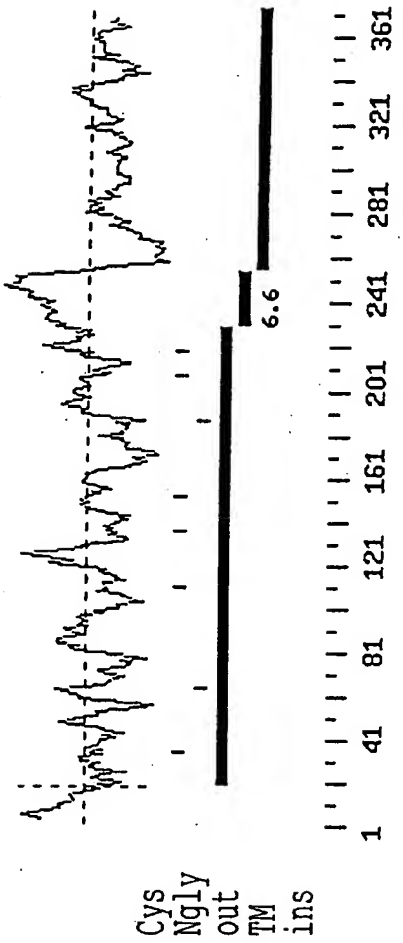

Fig. 27